

**This Page Is Inserted by IFW Operations
and is not a part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- **BLACK BORDERS**
- **TEXT CUT OFF AT TOP, BOTTOM OR SIDES**
- **FADED TEXT**
- **ILLEGIBLE TEXT**
- **SKEWED/SLANTED IMAGES**
- **COLORED PHOTOS**
- **BLACK OR VERY BLACK AND WHITE DARK PHOTOS**
- **GRAY SCALE DOCUMENTS**

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

THIS PAGE BLANK (USPTO,

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:25:54 ; Search time 22.7967 Seconds
(without alignments)
5267.077 Million cell updates/sec

Title: SEQ4-5-25-12
Perfect score: 6774
Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPYGLDVEV 1249

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 73:*
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6672	98.5	1255	1 A24571	protein-tyrosine k
2	5861.5	86.5	1254	2 I48161	p-185 precursor -
3	5854	86.4	1260	1 TVRINU	protein-tyrosine k
4	3135	46.3	1210	1 GQHUE	epidermal growth f
5	3114	46.0	1210	2 A53183	epidermal growth f
6	3085.5	45.5	1223	1 TVCHLV	epidermal growth f
7	2968.5	43.8	1308	2 A47253	protein-tyrosine k
8	2680	39.6	1166	1 S06142	kinase-related tra
9	2414	35.6	1342	2 A36223	epidermal growth f
10	2333.5	34.4	1339	2 JC4387	protein-tyrosine k
11	1766.5	26.1	698	1 TVFVLV	protein-tyrosine k
12	1703	25.1	604	1 TVYUHV	epidermal growth f
13	1652.5	24.4	1330	1 GQFPE	protein-tyrosine k
14	1647	24.3	544	2 S35745	kinase-related tra
15	1640	24.2	545	2 S00727	protein-tyrosine k
16	1623	24.0	540	2 B44776	protein-tyrosine k
17	1621	23.9	540	1 TVFVEB	epidermal growth f
18	1506	22.2	644	2 A63225	protein let-23 [lm
19	1300.5	19.2	1323	2 B88257	protein-tyrosine k
20	1300.5	19.2	1374	2 S70712	protein-tyrosine k
21	1190.5	17.6	1369	2 S70712	epidermal growth f
22	1182	17.4	1717	1 A45558	epidermal growth f
23	1117	16.5	527	2 A42032	epidermal growth f
24	949.5	14.0	843	2 A27131	protein-tyrosine k
25	806.5	11.9	346	2 S13807	protein-tyrosine k
26	754.5	11.1	311	2 S13808	insulin-like growth
27	722	10.7	1363	2 T43220	insulin receptor p
28	698.5	10.3	1382	1 INHUR	insulin-like growth
29	694.5	10.3	1607	2 T43212	

ALIGNMENTS

RESULT 1

A24571
protein-tyrosine kinase (EC 2.7.1.112) erbB2 precursor - human
N:Alternate names: c-erb-B-2 protein precursor; Kinase-related transforming protein erb
C:Species: Homo sapiens (man)
C:Date: 25-Oct-1987, #sequence revision 06-Dec-1996 #text change 11-Jun-1999
C:Accession: A24571; A25491; A44188; B44188; I59509; I57622
R:Yamamoto, T.; Ikawa, S.; Akiyama, T.; Semba, K.; Nomura, N.; Miyajima, N.; Saito, T.;
Nature 319, 230-234, 1986
A:Title: Similarity of protein encoded by the human c-erb-B-2 gene to epidermal growth
A:Reference number: A24571; MUID:86118663; PMID:3003577
A:Accession: A24571
A:Molecule type: mRNA
A:Residues: 1-1255 <VAM>
A:Cross-references: GB:X03363; NID:g31197; PIDN:CAA27060.1; PID:g31198
R:Semba, K.; Kanata, N.; Toyoshima, K.; Yamamoto, T.
Proc. Natl. Acad. Sci. U.S.A. 82, 6497-6501, 1985
A:Title: A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB-1/epider
A:Reference number: A25491; MUID:86016729; PMID:2995967
A:Accession: A25491
A:Molecule type: DNA
A:Residues: 737-1031 <SEM>
A:Cross-references: GB:M11767; NID:g182163; PIDN:AAA35808.1; PID:g553282
R:Cousens, L.; Yang-Feng, T.L.; Liao, Y.C.; Chen, E.; Gray, A.; McGrath, J.; Seeburg, J.
Science 230, 1132-1139, 1985
A:Title: Tyrosine kinase receptor with extensive homology to EGF receptor shares chromo
A:Reference number: A44188; MUID:86070181; PMID:2999974
A:Accession: A44188
A:Molecule type: DNA
A:Residues: 740-910 <COU1>
A:Cross-references: GB:M12036; NID:g183988; PIDN:AAA35978.1; PID:g183989
A:Accession: B44188
A:Molecule type: mRNA
A:Residues: 1-517, 'RALL', 522, 'S', 524-654, 'V', 656-1169, 'A', 1171-1255 <COU2>
A:Cross-references: GB:M11730; NID:g183986
R:King, C.R.; Kraus, M.H.; Aaronson, S.A.
Science 229, 974-976, 1985
A:Title: Amplification of a novel v-erbB-related gene in a human mammary carcinoma.
A:Reference number: I59509; MUID:85272597; PMID:2992089
A:Accession: I59509
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 832-909 <REX>
A:Cross-references: GB:I29395; NID:g459807; PIDN:AAA35809.1; PID:g459808
R:Tal, M.; King, C.R.; Kraus, M.H.; Ullrich, A.; Schlessinger, J.; Givol, D.
Mol. Cell. Biol. 7, 2597-2601, 1987
A:Title: Human HER2 (neu) promoter: evidence for multiple mechanisms for transcriptional
A:Reference number: I57622; MUID:87286898; PMID:3039351
A:Accession: I57622
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-191 <TAL>

insulin receptor p
insulin receptor p
protein-tyrosine k
insulin receptor-r
insulin receptor-r
insulin-like growth
insulin-like growth
insulin receptor -
insulin receptor -
protein-tyrosine k
protein-tyrosine k
tyrosine kinase Mp
protein-tyrosine k
protein-tyrosine k
protein-tyrosine k

A;Cross-references: GB:M16792; NID:g183983; PIDN:AAA58637.1; PID:g553332
 C;Comment: Amplification and overexpression of this erbB-related gene occurs in about 30
 C;Genetics:
 A;Gene: GDB:ERBB2; NCL; NEU; HER-2
 A;Map position: 17q21.1-17q21.1
 A;Introns: 25/1; 75/3; 147/1; 883/3
 A;Note: the list of introns is incomplete
 C;Function:
 A;Description: catalyzes the phosphorylation of a peptidyl tyrosine residue by ATP
 C;Superfamily: epidermal growth factor receptor; protein kinase homology
 C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
 inase
 F;1-21/Domain: signal sequence #status predicted <SIG>
 F;22-1255/Product: protein-tyrosine kinase erbB2 #status predicted <MAT>
 F;22-653/Domain: extracellular #status predicted <EXT>
 F;70-304/Domain: EGF receptor extracellular domain repeat <EB1>
 F;395-605/Domain: EGF receptor extracellular domain repeat <EB2>
 F;654-675/Domain: transmembrane #status predicted <TM>
 F;676-1255/Domain: intracellular #status predicted <INT>
 F;718-983/Domain: intracellular #status predicted <KIN>
 F;726-734/Region: protein kinase ATP-binding motif
 F;68,124,187,259,530,571,623/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F;686/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
 F;753/Active site: Lys #status predicted
 F;1139,1221,1222,1248/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation)

Query Match 98.58; Score 6672; DB 1; Length 1255;
 Best Local Similarity 98.24; Pred. No. 2.2e-265;
 Matches 1232; Conservative 5; Mismatches 12; Indels 6; Gaps 1;

QY 1 MELAALCRWGLLLALLPPGAASQTQCT-----QYIKANSKFIGITELLYQGCVVQGNL 54
 DB 1 MELAALCRWGLLLALLPPGAASQTQCTGCTDMKRLPASPTHLDMLHLHYQGCVVQGNL 60

QY 55 ELTYLPTNASLFLQDIQEVQGVYLIHAHQVQVPLQRLIRVGTQLFEDNYALAVLDNG 114
 DB 61 ELTYLPTNASLFLQDIQEVQGVYLIHAHQVQVPLQRLIRVGTQLFEDNYALAVLDNG 120

QY 115 DPLNNTPTVGTGSPGLRELQRLSLTEILKGGVLIQBNPOLCYQDTILWKDIFHKNOLA 174
 DB 121 DPLNNTPTVGTGSPGLRELQRLSLTEILKGGVLIQBNPOLCYQDTILWKDIFHKNOLA 180

QY 175 LTLLDINRSRACHPCSPMKSGSCWGESSEDCOSLRTVTCAGGCARCKGKPLPTDCCHEQC 234
 DB 181 LTLLDINRSRACHPCSPMKSGSCWGESSEDCOSLRTVTCAGGCARCKGKPLPTDCCHEQC 240

QY 235 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRTYTFGASCVTACP 294
 DB 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRTYTFGASCVTACP 300

QY 295 YNYLSTDVGSCTLVCPHMQEVAEDGTORCEKSPCARVCYGLGMEHLREVRATSAN 354
 DB 301 YNYLSTDVGSCTLVCPHMQEVAEDGTORCEKSPCARVCYGLGMEHLREVRATSAN 360

QY 355 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLOEQLOVFTELEETIGVLIYISAWPDSLP 414
 DB 361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLOEQLOVFTELEETIGVLIYISAWPDSLP 420

QY 415 DLVSFQNLQVIRGRIILHNGAYSITLQGLGSIWGLSLRELGSGLAIHHNTHLCFVHTV 474
 DB 421 DLVSFQNLQVIRGRIILHNGAYSITLQGLGSIWGLSLRELGSGLAIHHNTHLCFVHTV 480

QY 475 PWDOLFERNHQAALLHTANRDECEVGEGLACHOLCARGHCWGPGPTQCNCSQFLRGQC 534
 DB 481 PWDOLFERNHQAALLHTANRDECEVGEGLACHOLCARGHCWGPGPTQCNCSQFLRGQC 540

QY 535 VEECRVLQGLPREYVNRHCLCPHPECQPNQSVTCFGEADOCVCAHYKDPFFCVARC 594
 DB 541 VEECRVLQGLPREYVNRHCLCPHPECQPNQSVTCFGEADOCVCAHYKDPFFCVARC 600

QY 595 PGVKPDLSPYMPIWKPDPDEGACQPCPINCTHSCVDLDDKGCPEAQASPLTSISAVVG 654

RESULT 2

I48161
 p-185 precursor - golden hamster
 C;Species: Mesocricetus auratus (golden hamster)
 C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 18-Jun-1999
 C;Accession: I48161
 R;Nakamura, T.; Ushijima, T.; Ishizaka, Y.; Nagao, M.; Arai, M.; Yamazaki, Y.; Ishikawa, Y.
 Gene 140 251-255, 1994
 A;Title: Cloning and activation of the Syrian hamster neu proto-oncogene.
 A;Reference number: I48161; MUID:94193007; PMID:7908275
 A;Accession: I48161
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: mRNA
 A;Residues: 1-1254 <RES>
 A;Cross-references: GB:D16295; NID:g493236; PIDN:BAA03801.1; PID:g747595
 C;Genetics:
 A;Gene: neu
 C;Superfamily: epidermal growth factor receptor; protein kinase homology
 C;Keywords: ATP
 F;718-983/Domain: protein kinase homology <KIN>
 F;726-734/Region: protein kinase ATP-binding motif

Query Match 86.5%; Score 5861.5; DB 2; Length 1254;
 Best Local Similarity 86.1%; Pred. No. 2.6e-232;
 Matches 1081; Conservative 59; Mismatches 108; Indels 7; Gaps 2;

QY 1 MELAALCRWGLLLALLPPGAASQTQCT-----QYIKANSKFIGITELLYQGCVVQGNL 54
 DB 1 MELAALCRWGLLLALLSPGASGTQCTGCTDMKRLPASPTHLDIVRHLYQGCVVQGNL 60

QY 55 ELTYLPTNASTLSFLQDIOEVQGVVLTAHNOVROVPLQRLIRVGTOLFEDNYALAVLDNG 114
Db 61 ELTYLPANATLSFLQDIOEVQGVYMLIAHSQVRHVPLOQLRIRVGTOLFEDKYALAVLDNR 120
QY 115 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQORNPCOLCYQDTILWKDIFHKKNOLA 174
Db 121 DPLDNVTTATGRTPEGLRELQRLSLTEILKGGVLIIRGNPCOLCYQDTILWKDVFRKKNOLA 180
QY 175 LTLIDTNRSRACHPCSPMCKSGRCSGESSDCQSLTRTVCCAGCARCKGFLPTDCCHEQC 234
Db 181 PVDIDTNRSRACPPCAPACKDNHCWGPEDCQTLTGTTIAPRAVPAARARLPTDCCHEQC 240
QY 235 AAGCTGPKGSDCLACHFNHSGICELHCPALVTYNTDTFESMNPGRYTFGASCVTAC 294
Db 241 AAGCTGPKGSDCLACHFNHSGICELHCPALVTYNTDTFESMNPGRYTFGASCVTTC 300
QY 295 YNYLSTDVGSCTLVCPILHNOEVTAEQDTCRCKSKPCARVCYGLGMEHLREVRVATSN 354
Db 301 YNYLSTEVGCTLVCPILHNOEVTAEQDTCRCKSKPCARVCYGLGMEHLRGARAITSN 360
QY 355 IQEFAGCKIFGSLAFIPESFDGDPASNTAPLOEQVFTLEETIGLYISAWPDSL 414
Db 361 IQEFAGCKIFGSLAFIPESFDGNPSSGIAPLTPEQLOVFTLEETIGLYISAWPDSL 420
QY 415 DLSVFQNLQVIRGRILHNGAYSITLQGLGISWGLRSLRELGLSLGLIHHNTHLCFVHTV 474
Db 421 DLSVFQNLVIRGRVLDHNGAYSITLQGLGISWGLRSLRELGLSLGLIHHNTHLCFVHTV 480
QY 475 PDOLFERNHOALHTANRDEDCVGEGLACHOLCARGHCWGPQTCVNCOSFLRGQC 534
Db 481 PDOLFERNHOALHTANRDEDCVGEGLACHOLCARGHCWGPQTCVNCOSFLRGQC 540
QY 535 VBCRYLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFVCVARC 594
Db 541 VKECRWKGLPREYVNGKCLPCHPECPQNSVTCFGEADQCTACPHYKDSPPFCVARC 600
QY 595 PSGVKPDLVMPYKWPDEBAGACQPCINETHSCVDLDKGCPEAQRASPLTSIVSAVVG 654
Db 601 PSGVKPDLVMPYKWPDEBAGACQPCINETHSCVDLDKGCPEAQRASPLTSIVSAVVG 660
QY 655 ILLAVVLGVVFGTLIKRQOKIRKYTMRLLOTELVEPLTPSGAMPNOAQRILKETEL 714
Db 661 ILLFLVGVVFGTLIKRQOKIRKYTMRLLOTELVEPLTPSGAMPNOAQRILKETEL 720
QY 715 RKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKILDEAYVMAGVGP 774
Db 721 RKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKILDEAYVMAGVGP 780
QY 775 YVSRLLGICLTSTVQLVTPQMLPYGCLLDHVNRGRGLSGODLLNWCWQIAKGMVLEDDR 834
Db 781 YVSRLLGICLTSTVQLVTPQMLPYGCLLDHVNRGRGLSGODLLNWCWQIAKGMVLEDDR 840
QY 835 LVHRDLAARNVLKSPNHVKITDPLGLARLIDDETEYHADGGKVPKWALESILRRRT 894
Db 841 LVHRDLAARNVLKSPNHVKITDPLGLARLIDDETEYHADGGKVPKWALESILRRRT 900
QY 895 HQSDVMSYGVVWELMTFGAKPYDGIIPAREIPDLEKGERLPPOPICTIDVYIMVKCWM 954
Db 901 HQSDVMSYGVVWELMTFGAKPYDGIIPAREIPDLEKGERLPPOPICTIDVYIMVKCWM 960
QY 955 IDSECRPRFELVSEFSRMARDPQRFVVIQNEQDLGPASPLDSTFYRSLDEDDMDGLVDA 1014
Db 961 IDSECRPRFELVSEFSRMARDPQRFVVIQNEQDLGPSSPLDSTFYRSLDEDDMDGLVDA 1020
QY 1015 EYLVLPQOQFFCPDPAPGAGWVHHRSSSTRSGGDLTLGLEPSEEAAPRSLAPSEG 1074
Db 1021 EYLVLPQOQFFCPDPAPGAGWVHHRSSSTRSGGDLTLGLEPSEEAAPRSLAPSEG 1080
QY 1075 AGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSBTGDGVAPLTCSPQPEYV 1134
Db 1081 AGSDVFEGLGMAATKGPQISPRDLSPLQRYSEDPTVPLPSBTGDGVAPLTCSPQPEYV 1140
QY 1135 NQPDVPRQPSPREGPLPAARPAAGATLERAKTSLSPGKNVGVKDVFAFGGAVENPEYLTQ 1194

Db 1141 NOPEVRPQPLTPEGLPEVRPAGATLERPKTSLSPCKGWKDVFTFGGAVENPEYLVPR 1200
QY 1195 GGAAPQPPHPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVVP 1249
Db 1201 GGSASQPH-PPALCPAFDNLVYWDQDPSERGSFPNTFEGTPTAENPEYLGLDVVP 1254
RESULT 3
TVRTNU
protein-tyrosine kinase (BC 2.7.1.112) neu precursor - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 11-Jun-1999
C:Accession: A24562; A61204
R:Bargmann, C.I.; Hung, M.C.; Weinberg, R.A.
Nature 319, 226-230, 1986
A:Title: The neu oncogene encodes an epidermal growth factor receptor-related protein.
A:Reference number: A24562; MUID:86118662; PMID:3945311
A:Accession: A24562
A:Molecule type: mRNA
A:Residues: 1-1260 <BAR>
A:Cross-references: EMBL:X03362; NID:g56745; PIDN:CAA27059.1; PID:g56746
R:Masui, T.; Mann, A.M.; Macatee, T.L.; Garland, E.M.; Okamura, T.; Smith, R.A.; Cohen, Carcinogenesis 12, 1975-1978, 1991
A:Title: Direct DNA sequencing of the rat neu oncogene transmembrane domain reveals no 2-thiazolylformamide or N-methyl-N-nitrosourea.
A:Reference number: A61204; MUID:92035293; PMID:1682063
A:Accession: A61204
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 637-663, 'V', 665-702 <MAS>
A>Note: authors translated the codon GCA for residue 25 as Val
C:Genetics:
A:Gene: neu
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
F:1-13/Domain: signal sequence #status predicted <SIG>
F:20-1260/Product: protein-tyrosine kinase neu #status predicted <MAT>
F:658-680/Domain: transmembrane #status predicted <TMN>
F:723-988/Domain: protein kinase homology <KIN>
F:731-739/Region: protein kinase ATP-binding motif
F:71,191,263,535,576,634/Binding site: carboxydrate (Asn) (covalent) #status predicted
F:691/Binding site: phosphate (Thr) (covalent) #status predicted
F:758/Active site: Lys #status predicted
F:882,1227,1253/Binding site: phosphate (Tyr) (covalent) #status predicted
Query Match 86.4%; Score 5854; DB 1; Length 1260;
Best Local Similarity 86.1%; Pred. No. 5.4e-232;
Matches 1082; Conservative 54; Mismatches 113; Indels 8; Gaps 3;
QY 1 MELALCRWGLLLALLPPGAASVQCT-----QVIAKSKFIFGITELLYQGVVQGNL 54
Db 4 MELANCRWGLLLALLPPGIAGTCTGTDMKRLRPLASPETHLDMRLHYQGVVQGNL 63
QY 55 ELTYLPTNASTLSFLQDIOEVQGVYMLIAHNOVROVPLQRLIRVGTOLFEDNYALAVLDNG 114
Db 61 ELTYLPANATLSFLQDIOEVQGVYMLIAHNOVROVPLQRLIRVGTOLFEDKYALAVLDNR 120
QY 115 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQORNPCOLCYQDTILWKDIFHKKNOL 173
Db 121 DPLDNVTTATGRTPEGLRELQRLSLTEILKGGVLIIRGNPCOLCYQDTILWKDVFRKKNOL 183
QY 175 LTLIDTNRSRACHPCSPMCKSGRCSGESSDCQSLTRTVCCAGCARCKGFLPTDCCHEQC 233
Db 181 PVDIDTNRSRACPPCAPACKDNHCWGPEDCQTLTGTTIAPRAVPAARARLPTDCCHEQC 243
QY 235 AAGCTGPKGSDCLACHFNHSGICELHCPALVTYNTDTFESMNPGRYTFGASCVTAC 293
Db 241 AAGCTGPKGSDCLACHFNHSGICELHCPALVTYNTDTFESMNPGRYTFGASCVTTC 303
QY 295 YNYLSTDVGSCTLVCPILHNOEVTAEQDTCRCKSKPCARVCYGLGMEHLREVRVATSA 353
Db 301 YNYLSTEVGCTLVCPILHNOEVTAEQDTCRCKSKPCARVCYGLGMEHLRGARAITSD 363

A:Molecule type: mRNA
 A:Residues: 969-971, 'K', 973-1115, 'D' <EIS>
 A:Cross-references: EMBL:Z12608
 R:Heisermann, G.J.; Gill, G.N.
 J. Biol. Chem. 263, 13152-13158, 1988
 A:Title: Epidermal growth factor receptor threonine and serine residues phosphorylated
 A:Reference number: A28941; MUID:88330814; PMID:3138233
 A:Accession: A28941
 A:Molecule type: protein
 A:Residues: 689-694, 'X', 696-704, 'L', 706-707, 989-992, 'XX', 995-996, 'X', 998-1000, 1002-1009, R:Ribos, M.L.; Dunn, A.R.; Alexander, W.S.
 submitted to the EMBL Data Library, April 1994
 A:Description: The complete cDNA sequence of the Mouse Epidermal Growth Factor Receptor
 A:Reference number: S45325
 A:Accession: S45325
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-971, 'K', 973-1210 <VER>
 A:Cross-references: EMBL:X78987; NID:G488830; PIDN:CAA55587.1; PID:G488831
 R:Paria, B.C.; Das, S.K.; Andrews, G.K.; Dev, S.K.
 Proc. Natl. Acad. Sci. U.S.A. 90, 55-59, 1993
 A:Title: Expression of the epidermal growth factor receptor gene is regulated in mouse b
 A:Reference number: 149643; MUID:93126380; PMID:7678348
 A:Accession: 149643
 A>Status: translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 12-20, 132 <RES>
 A:Cross-references: GB:L06864; NID:G193001; PIDN:AAA3029.1; PID:9567201
 C:Genetics:
 C:Gene: EGFR
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: ATP; growth factor receptor; kinase-related transforming protein; phosphop
 F:1-24/Domain: signal sequence #status predicted <SIG>
 F:548-670/Domain: transmembrane #status predicted <TM>
 F:712-977/Domain: protein kinase homology <KIN>
 F:720-728/Region: protein kinase ATP-binding motif
 F:680,695/Binding site: phosphate (Thr) (covalent) #status experimental
 F:697,1070,1071/Binding site: phosphate (Ser) (covalent) #status experimental
 F:993/Binding site: (or 997) phosphate (Ser) (covalent) #status experimental
 F:1028/Binding site: (or 1030 or 1032) phosphate (Ser) (covalent) #status experimental
 F:1197/Binding site: phosphate (Tyr) (covalent) #status experimental

Query Match 46.0%; Score 3114; DB 2; Length 1210;
 Best Local Similarity 49.5%; Pred. No. 3, 8e-120;
 Matches 630; Conservative 170; Mismatches 355; Indels 118; Gaps. 25;
 QY 11 LLALLPPGAA--STOVC-----TQYTKANSKFTGITEILYQGCQVQVQGNLEUTYLP 61
 DB 14 LLTALCAAGGAALEKKVKCGQTSNRLTQGLTFEDHFLSL-QRMYNCEVVLGNLEITYVQR 72
 QY 62 NASLSPLDIOEQVGVYLTAHNOVROVPLORLRIVRGTOLFEDNYALAVLDNGDPLNNTT 121
 DB 73 NYDLSPFKTIQEVAGVYLTAHNTVERIPLENQIRGNALYENTYALALSN----- 124
 QY 122 PVTGASPGGLREILQLRSLTEILKGGVLIQRNPOLCYQDTILWKDI----FHKNNQALATL 177
 DB 125 --YGNRTGLRELPMNLQELLIGAVRFNNPILCNWDIIQRDIQVNFMSNMDL--- 180
 QY 178 IDNRSRACHPCSPMKCGRSGWESSEDCOSLTRVCAGGCA-RCKGPLPTDCCHBQCAA 236
 DB 181 --QSHPSSCPCKDPCSPNGSCWGGGEEQKLTIIICAOQCSHRRCGRSPSCCHNOCAA 238
 QY 237 GCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTGASCVTACPN 296
 DB 239 GCTGPRESCLVCQKQFQDEATCKOTCPPLMLNPTTYQMDVNPPEGKYSFGATCVKCKPN 298
 QY 297 YLSTDVGSCTVLCPLHNQVTAEDGTQRCCKSKPCARVCYGLGMEHLREAVRATISNTQ 356
 DB 299 YVYTDHGSVCVRACGPDYEV-EDGIRKCKCKDGPCKVCNGIGEGFKDTLSINATNIK 357
 QY 357 EFAGCKKIFGSLAFLPESFDGDPASNTAPLOEQVFTLEITGLYLVISAWPDSLPDL 416
 DB 358 HFKYCTAISGDLHILPFAFKGDSFTRTPPLDPRELEILTKVKEITGFLLIQAWPDNWTDL 417

RESULT 6

TVCHLV

epidermal growth factor receptor precursor - chicken

N:Contains: protein-tyrosine kinase (EC 2.7.1.12) erbB

C:Species: Gallus gallus (chicken)

C>Date: 28-Feb-1986 #sequence_revision 05-May-1995 #text_change 04-Feb-2000

C:Accession: A27720; A00643

R:Jax, J.; Johnson, A.; Howk, R.; Sap, J.; Bellot, F.; Winkler, M.; Ullrich, A.; Vennet

Cell. Biol. 8, 1970-1978, 1988

A:Title: Chicken epidermal growth factor (EGF) receptor: cDNA cloning, expression in mou

A:Reference number: A27720; MUID:88261272; PMID:3260329

A:Accession: A27720
A:Molecule type: mRNA
A:Residues: 1-1223 <LAX>
A:Cross-references: GB:M20386
R:Nilsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M.
Cell 41, 719-726, 1985
A:Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro
A:Reference number: A00643; MUID:85228222; PMID:2988784
A:Accession: A00643
A:Molecule type: mRNA
A:Residues: 585-1223 <NIL>
A:Cross-references: GB:M10066
C:Genetics:
A:Gene: erbB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: alternative splicing; ATP; autophosphorylation; glycoprotein; growth factor
specific protein kinase
F:1-30/Domain: signal sequence #status predicted <SIG>
F:31-1223/Product: epidermal growth factor receptor #status predicted <MAT>
F:31-654/Domain: extracellular #status predicted <EXT>
F:81-307/Domain: EGF receptor extracellular domain repeat <EE1>
F:397-610/Domain: EGF receptor extracellular domain repeat <EE2>
F:655-677/Domain: transmembrane #status predicted <TM>
F:678-1223/Domain: intracellular #status predicted <INT>
F:719-984/Domain: protein kinase homology <KIN>
F:727-735/Region: protein kinase ATP-binding motif
F:136,202,280,361,370,422,575,580,615,635/Binding site: carbohydrate (Thr) (covalent) #s
F:192,650/Binding site: carbohydrate (Ser) (covalent) #status predicted
F:687/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F:754/Active site: Lys #status predicted
F:1100,1183,1208/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #stat
Query Match 45.5%; Score 3085.5; DB 1; Length 1223;
Best Local Similarity 48.4%; Pred. No. 5,6e-119;
Matches 628; Conservative 175; Mismatches 342; Indels 153; Gaps 27;
QY 8 RWGLLALLPPGAA-----STQVC-----TOYIKANSKFIGITELYQCGVQGNL 54
DB 13 RGAALVLLLLGVALCSAVEEKKVCGQTNKKLTQLGHVEDHFTSL-QRMNNCEVLSNL 71
QY 55 ELTYPLTNASLFLQDIQVQGVLTIAHNVQVQLRIVRGTQLFEDNVALAVLDNG 114
DB 72 EITYVHNHDLFLTKIQEVAGVLTIALNMVDVPIENQIIRGNVLYDNSFALAVLSY 131
QY 115 DPLNNTPTVTGASPGRLQLRLSLTEILKGGVLIORNPOLCVQDTILWKDIFHKNNOLA 174
DB 132 H-MNKTQ-----GLRELPMKRLSEILNGGVKLSNPKLCNMTVLNNDLIDTSRK-P 181
QY 175 LFLID-TNRSRACHPCSPCKSGRWBSSEDCQSLTRTVACGGA-RCKGPLPTDCHE 232
DB 182 LTVLDFASNLSSCPKCHPNCTEDHCWAGEQNCQTLTKVICAQCSGRCKGVPSDCCHN 241
QY 233 QCAAGCTGPKHSDCLACLFHNSIGCELHCPALVTYNTDTFESMPNPEGRYTFGASCVTA 292
DB 242 QCAAGCTGPRSDCLACRFDRDATCKDTPPLVLYNPTYQMDVNPPEKYSFGATCVRE 301
QY 293 CPYNYLSTDVGSCTLVCPLHNQEVTAEDGTQCECKSPCARVCYGLGMHLREYRAVTS 352
DB 302 CPHYVYVDHGSVRSNCTDTEV-EENGVRKCKKCDGLCKVKNVNGIGIGELKGLISINA 360
QY 353 ANIQEFAGCKKIFGSLAFIPESFDGDPASNTAPLOEQLOVFEITGLYLYISAWPDS 412
DB 361 TNIDSPKNTKINGDVSILPVAFLGDAFTKTLPLDPKLDVFTVKEISGFLLIQAWPDN 420
QY 413 LPDLSVFQNLQVIRGRIHNGAYSITLQGLGTSWGLSLRLGSLGLALIHNTLCFVH 472
DB 421 ATDLYAFENLEIRGTQKHQGYSLAVNLKIQSLGLSLKLSKISDGIATIMKNKRLCYAD 480
QY 473 TVPWQDLFRNPQALLHTANRPEDECVGSLACHOLCARGHCWGPGTQVCNCSFLGQ 532
DB 481 TNWNSLFAEQSKTKIQRNKNKDTRDHRVCDPLCSVDGCGWGPCHFCRFFSRQK 540
QY 533 ECVEECRVLQGLPREYVNAHRLCPHPECQPNQ---SVTCFGPEADQCACAHYKDPFF 589

RESULT 7

A47253
epidermal growth factor receptor, HER4 - human
C:Species: Homo sapiens (man)
C:Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 18-Jun-1999
C:Accession: A47253
R:Flowerman, G.D.; Culouscou, J.M.; Whitney, G.S.; Green, J.M.; Carlton, G.W.; Foy, L.; Nee
Proc. Natl. Acad. Sci. U.S.A. 90, 1746-1750, 1993
A:Title: Ligand-specific activation of HER4/p180erbB4, a fourth member of the epidermal
A:Reference number: A47253; MUID:93189574; PMID:8383326
A:Accession: A47253
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1-1308 <PLO>
A:Cross-references: GB:L07868; NID:g337359; PIDN:AAB59446.1; PID:g337360
A>Note: sequence extracted from NCBI backbone (NCBIP:126842)
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; growth factor receptor
F:716-981/Domain: protein kinase homology <KIN>
F:724-732/Region: protein kinase ATP-binding motif

DB 541 ECVKQCNILQGEPRFERDSEKCLPCHSECLVQNSTAYNTTCSPGPDHCKMAHFDGPH 600
QY 590 CVARCPGKVPDLNWKFPDEGACQPCPINCTHSCVDLDDKGCAPAEQASPLTSIV 649
DB 601 CVKACPAVGLGENDTL-VKAYADANAVCOLCHPNCTRGCKGPLEGCP---NGSKTPSTA 656
QY 650 SAVV-GILLVVVLGVVFGILKRRQOKIRKYTMRLLQSTELVELFPLTSGAMPNOQMRI 708
DB 657 AGVWGLLCLVWVGLGIGLYLRRR-HIVKRTLRLLQERELVEPLTPSGEAPNQAHLAI 715
QY 709 LKETELRKVKVLGSCAFGTIVYKGIWIPGENVKIPIVAIKVLRENTSPKANKILDEAYYM 768
DB 716 LKETEFKVKVLGSGAFGTIVYKGLWIPGEKVKIPIVAIKELREATSPKANKILDEAYYM 775
QY 769 AGVGPVYSRLIGLICLTSTVOLVTQMPYVCLLDHVRNRRGLSGQDILLNWCQIAKMS 828
DB 776 ASVDNPHVCRLLIGLICLTSTVOLITQMPYVCLLDHVRNRRGLSGQDILLNWCQIAKMS 835
QY 829 YLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMALESI 888
DB 836 YLEERLVRDLAARNVLKTPQHVKITDFGLAKLGADEKEYHAEGKVPKMALESI 895
QY 889 LRRRTHQSDVWSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPOPPICITDVYMI 948
DB 896 LHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIPIASEISSVLEKGERLPOPPICITDVYMI 955
QY 949 MYKCMWIDSECRPRELVSFRSFRWARDPQRFVLIQ-NEDLGPASPLDSTFVRSLLDDDD 1007
DB 956 MYKCMWIDADSRPRELIAEISFRWARDPRLVYIQTQGERMHLPSPTDSKPYRTLMEEED 1015
QY 1008 MGDVLDABEYLVPOQGFPCPAPAGAGGVHHRHRSSTRSGGDLTLGLLEPSEBEAPRS 1067
DB 1016 MEDIVDADEYLVPHGQFF-----NSPST-----SRT 1041
QY 1068 PL-----APSEGAGSDVDFDGLGMGAAGKLSLPTHTDPSLPQRYSEDPTVPLPSET--DG 1120
DB 1042 PLLSLSLSATSNNSATNCID-----RNGQCHPVREDSFVQRYSSDPTGNFLESSIDDG 1093
QY 1121 VYAPLTCSPQEVYVNPQDVRPQPPSPREGPLPAARPAATLERAKTLSPGKNGVVKDVP- 1179
DB 1094 FL-----PAPEYVQ---LMPKKPS-----TAMVOIQYNNISL 1124
QY 1180 -----AFGGAVENPEYLTPOGGAAPQHPHPPAPFAPAFDNLVYWDQ----- 1219
DB 1125 TAISKLPWDSRYQNSHSTAVDNPEYL-----NTNOSPLAKTVFESSPYIQSGNHQI 1176
QY 1220 --DPPE-----FGAPPSTFKGTPTAENPEYLGLDVP 1248
DB 1177 NLDNDPYQQDFLNPETKPNGLLKVPAENPEYLRVAAP 1214

Query Match		43.8%	Score 2968.5;	DB 2;	Length 1308;
Best Local Similarity		45.1%	Pred. No. 3.6e-114;		
Matches		609;	Conservative 186;	Mismatches 369;	Indels 185; Gaps 30;
QY	9	WGLLLALLPPGAA-----STQVCTQYIKANSKFGITEL-----	LYQGVQVVOGNLE	55	
DB	8	WVWSLLVAAGTVQPSDSOSVCA--GTEKLSLSLDLEQQVRAIRKYYENCEVWNGLE	64		
QY	56	LTLYPTNASLSFLQDIQEOGVYLIHAHNOVROVPLQRLRIVRGTOLEFEDNYALAVLDNGD	115		
DB	65	ITSIEHRDLDFLRSVREVTGVYLVALNQRYLPLENLRIIRTKLYEDRYALAIPLNVR	124		
QY	116	PLNNTPTVTGASPGGLREQLRSLTEILKGGVLIQRNPOLCYQDITILWKDIFHKNNQAL	175		
DB	125	KDGNF-----GLQELGLKNLTILNGVGVVDQNKFLCYADTIHWQDIVRNWPSNL	175		
QY	176	TLIDTNRBRACHPCSPMKSGSRGESSDDCQLTRTYVCAGGC-ARCKGPLPTDCHEQC	234		
DB	176	TLVSTNGSGGCRCHKSCGTC-RWGPTENHCQTLTRVCAEQDCGRCYGPIYSDCCCHREC	234		
QY	235	AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPGRYTFGASCVTACP	294		
DB	235	AGGCSGPKDTDFACWNFNDGACVTCQCTQTFVYNPTTFQLEHFNAKITYGAFCKVCKP	294		
QY	295	YNYLSTDVGSCTLVCPHNNQEVTAEDGTQRCCKSKPCARVCYCGLGMHLEHREVRVTSAN	354		
DB	295	HNFV-VDSSSCVACRACSSKMEV-EENGIRMKCPKPTDIPCACDGICTGSLMSAQTVDSN	352		
QY	355	IOEFACKKIFGSLAFPLPSFDGDPASNTAPLOPEQLQVETLEETGYLYISAMPDSLP	414		
DB	353	IDKFINCTINGNLIFLVGIIHGDYPNAIEADPEKLNIVFRTVREITGFLNQSPPNMT	412		
QY	415	DLVSFONLOVIRGRIHLNGAYSILTLOGLGISMLGRSLRELGLSLIHHNTHLCFVHTV	474		
DB	413	DSVFSNLVTIGRVLVYSGLSLLILKQOGLTSLOFSLKEISAGNYITDNLKYYHTI	472		
QY	475	PMDQLFRNPQALLHTANRPEDCVGEGLACHQLCARGHCWGPQTQVCNCSQFLRGQRC	534		
DB	473	NMTTLFSTINQIRIVIRDRKAENCTAEGVMCNHLCSDDCWGPGDPDCLSCRRFSRGIC	532		
QY	535	VRECRVLQPLREYNARHCLPCHPECQ-ONGSVTCFGEADQCACAHYKDPDFCVAR	593		
DB	533	IESCNLYDGEFEFENGSTICVECDQCEKXEDGLLTCHGFGPDNCTKCSHFKDGPNCEK	592		
QY	594	CPSGVKPDLSPYPIWPKFPEEGACQPCINCTHSCVDLDDKGC-----PAEQRA	642		
DB	593	CPDGLQANSF--IPKADPDRECHPCNCTQCGNCTSHDCIYYPWTGHTLQOAR-	649		
QY	643	SPLTSTVSAVV-GILVVVVLGVVFGILIKRROQKIRKYTMRLLOTELVEPLTPSGAMP	701		
DB	650	TPL--IAAGVIGLFTLVIVGLTFVAVVRRKSII-KKRALRRFL-ETELVEPLTPSGTAP	705		
QY	702	NOAQRILKETELRKVKVLSGAFGVYKGIWIPDGENVKIPVAIKVLRENTSPKANKI	761		
DB	706	NOAQRILKETELRKVKVLSGAFGVYKGIWIPDGETVKIPVAIKILNETTGPKANVF	765		
QY	762	LDEAVVMAGVGPYSRLIGICTSTVQLVLTOLMPYGCILLDHVRENRGRGLSQDILLNWC	821		
DB	766	MEALIMASNDPHILVRLVGLVCLSPSTIQLVTQMPHGCULEYVHEKONIGSQLLNWC	825		
QY	822	QIAGKMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDDIDETEHADGGKVPFK	881		
DB	826	QIAGKMWLEERLVRHDLAARNVLKSPNHVKITDFGLARLLLEGEDEKEYNADGGKVPK	885		
QY	882	WMALESILRRRTHQSDVWSYGVVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPPIC	941		
DB	886	WMALECIHYRKTTHQSDVWSYGVVWELMTFGKPYDGIPTREIPDLLEKGERLPQPPIC	945		
QY	942	TIDVYIMVKWMDSECRPRELVSFESRMARDPQRFVWIONED-LGPASPLDSTFYR	1000		
DB	946	TIDVYVMVKWMDADSPKFKELAAEFSSRMARDPQRYLVIQGDDRMKLPSPNDSKFFQ	1005		
QY	1001	SLEDDDDMGDLVDAEYLVPOQGFCTPDPAAGAGGMVHRRHSSSTRSGGGDLTLGLEPS	1060		

DB	1006	NLLDEEDLEMDAEYLVP-QAFNIPPP-----IYTSRARI DNSRS-----EIGHSP	1053		
QY	1061	EEAPRS-----PLAP-SEGASDVDFDGLGWAAGK	1091		
DB	1054	PAYTMSGNQFYVRDGGFAAEQGVSVYRAPRTSTIPEAPVAQGATAEIPDDSCCNGTLRK	1113		
QY	1092	LOSLEPHTDPSPLQRYSEDPTVPLPS-----ETDGYVAPLTCSPOPEYVNOQDVRPQPP	1144		
DB	1114	PVAPHVQSDSSTQRIASADPTTFAPERSPRGELDEGYMTFMRDKPKQEYLVNPE-----	1167		
QY	1145	SPREGLPAAPAGATLERAKTLPSPKNGVVKDVFAGGAVENPEYLPQGGAAPOPHPP	1204		
DB	1168	---ENPFVSR---KNGDLQ-----ALDNPESHNASG-----PP	1196		
QY	1205	PA-----FSPAFDNLVYWDQDPPERGA--PP	1228		
DB	1197	KAEDYVNEPLYNLNTFANTLKAELYKNILSMPEKAKAFDNPDMYWNHSLPPRSTLQHP	1256		
QY	1229	STFKGTPT-----AENPEYL	1243		
DB	1257	DYLOEYSTKYFYKQNGRIRPIVAENPEYL	1285		
RESULT 8					
S06142					
protein-tyrosine kinase (EC 2.7.1.112) mrk-Y precursor - southern platyfish					
N/Alternate names: epidermal growth factor receptor homolog; kinase-related transformin					
C/Species: Xiphophorus maculatus (southern platyfish)					
C/Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text_change 18-Feb-2000					
C/Accession: S06142; S13809					
R/Wittbrodt, J.; Adam, D.; Malitschek, B.; Maeueller, W.; Raulf, F.; Telling, A.; Robert					
Nature 341, 415-421, 1989					
A/Title: Novel putative receptor tyrosine kinase encoded by the melanoma-inducing Tu lo					
A/Reference number: S06142; MUID:90015140; PMID:2797166					
A/Accession: S06142					
A/Molecule type: DNA					
A/Residues: 1-1166 <WIT>					
A/Cross-references: EMBL:X16891; NID:G65290; PIDN:CAA34770.1; PID:G65291					
R/Adam, D.; Maeueller, W.; Scharlt, M.					
Oncogene 6, 73-80, 1991					
A/Title: Transcriptional activation of the melanoma inducing Xmrk oncogene in Xiphophori					
A/Reference number: S13807; MUID:91125882; PMID:1846957					
A/Accession: S13809					
A/Status: preliminary; translation not shown					
A/Molecule type: DNA					
A/Residues: 821-1025, 'N', 1027-1098, 'A', 1100-1166 <ADA>					
A/Cross-references: EMBL:X56319; NID:G65284; PIDN:CAA39763.1; PID:G65285					
C/Genetics:					
A/Map position: Y					
A/Introns: 872/3; 898/1; 947/1; 979/3; 1025/3; 1056/1					
C/Superfamily: epidermal growth factor receptor; protein kinase homology					
C/Keywords: ATP; growth factor receptor; phosphotransferase; transmembrane protein; tyr					
F.1-25/Domain: signal sequence #status predicted <SIG>					
F.26-1166/Product: kinase-related transforming protein (Tu) #status predicted <MAT>					
F.707-972/Domain: protein kinase homology <KIN>					
F.715-723/Region: protein kinase ATP-binding motif					
Query Match					
Best Local Similarity 39.6%; Score 2680; DB 1; Length 1166;					
Matches 574; Conservative 163; Mismatches 388; Indels 140; Gaps 29;					
QY	4	AALCRWGLLLALLPPGAAS-----QVC-----TOYKANSKFGITELLYQGVQVVOGNLE	55		
DB	8	AALLQ--LLLVLSISRCCSTDPDRKVCQGTNQMTLMDNHYLKMKK-MYSGCNVYLENLE	64		
QY	56	LTLYPTNASLSFLQDIQEOGVYLIHAHNOVROVPLQRLRIVRGTOLEFEDNYALAVLDNGD	115		
DB	65	ITYTQENQDLDFLRSVREVTGVYLVALNQRYLPLENLRIIRTKLYEDRYALAIPLNVR	124		
QY	116	PLNNTPTVTGASPGGLREQLRSLTEILKGGVLIQRNPOLCYQDITILWKDIFHKNNQAL	175		

125	K-NPSSP--DVYQVGLKQLJLSNLTETLLSGGVKVSUHNPLLCNVETINWMDIVDTKSNPTM	181
176	TLIDTNRSRACHPCSPMKCKSGRCWGESSEDCQSLTRTVCAGGC-ARCKGPIPTDCCHBQC	234
182	NLI PHAFERQCQCKDHGCVNGSCWAPGPHGCQKFTKLLCASQCNRRCRGPXKPIDCCNEHC	241
235	AAGCTGPKHSDCLACHFNHSGI CELHCPALVTYNTDTFESMNPDEGRYTFGASCVTACP	294
242	AGGCTGPRATDCLACRDFNDGTCCTPPPKIYDIVSHQVVDVFNPKYTFEGAACVKECP	301
295	YNYLSTDVGSCCTLVCLPHNQBVTAEDGTORCEKSKPCARVCYGLGWEHLREVRVAVTSAN	354
302	SNYVYTE-GACVRSASGMLEVD--ENGRKSCKPCDGVCPKVCDDGJGSLSENTIAVNSTN	359
355	IQEFAGCKKIFGSLAFTPESGDGPASNTAPLOPEQLQVFFETLBEITGYLISAWPDSLP	414
360	IRSFENCTKINGDIILNRNPFEGDGHYKIGTMDPEHLWNLTIVKEITGYLVMWPNEMT	419
415	DLSVFQNLQVIRGRILHNHAYS-LTLQGLGISWLGRLSLRGLSGSLAIHHTHLCFVHT	477
420	SLSVFQNLLEIRGRITTSRGSFVYVQVRHLQWLGRLSKLSEVSGNVLKNTLQIRYANT	479
474	VPMDLFRNPHOALLHTANRPEDECVEGLACHQLCARGHCWGPQTCVNCQSOFLOGQE	533
480	INWRRLFRSEDQSTLEYDART-----ENQTCNNECEDGCW-PGPTNCVSLCHVDKGR	531
534	CVBECRYLQGLPREYVYVNAHCLPCHPBCQPONGSVTCFGEADOCVACAHYKDPFPCVAR	593
532	CVASCNLLOGEPRAOVDGRVCVQCHQECVLQVTDLSLTCYGGPGANCSAHPQDGPQILPR	591
594	CPSGVKPDLSYMPTWKFPDDEGACQPCPNCTHSCVDLDDKGCSPAQRASPLTISIVSAW	653
592	CPHGILGDGDTL-TWKYADKMGQCPQCHQNTQCGSPGLSGCRGD-IVSHSSLAUGLVS	649
654	GILLVVLGVVFGILLIKRROOKIRKYTWRRLLQETELVEPLTPSGAMPNQOAMRILKETE	713
650	GLLITIVALLIVLLRRRIK-RKRTIRCLLOEKELVEPLTPSGQAPNQAFRLKETE	708
714	LRKVKVLGSGAFGVYKGIWIPDCENVKIPVAIKVLRENTSPKANKELDEAYVMAGVS	773
709	FKDRVLGSGAFGVYKGLWNPDGENTIRIPVAIKVLRATSPKNOEVLDEAYVMASVDH	768
774	PYVSRLGICLTSTVQLVTQMPYGCILLDHVRENRLGSDOLLNWCWIAKMSYLEDV	833
769	PHVCRLLGICLTSAVLQVTQMPYGCLLDYVRQHOERICGWLLNWCWIAKGMNYLEER	828
834	RLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWWALESILRRFP	893
829	HLVHRDLAARNVLLKNPNHKITDFGLSKLTADAEKEYQADGGKVPKWWALESILOWTY	888
894	THQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPCTIDVYMWKCV	953
889	THQSDVMSYGVTVWELMTFGSKPYDGIIPAKIASVLENGERLPPOPCTTIEVYMIILKCV	948
954	MIDSECRPRELVSEFSRMARDPQRFVQIWNEDLGPASPLDSTFYRSLLEDDMGDLVD	1013
949	MIDPSSRPREFLVGEFSQMARDPARYLVIG--NLPSSLDRFLFSRLSSDD--DVVD	1003
1014	ABEYLVPOQGFCDPDPAAGWAGVHHRRSSSTRSGGDDLTLGLEPSEEAAPRPLAPSE	1073
1004	ADEYLLPYKRI-----NRQGS-----EPCIPPT	1026
1074	GAGSDVFDGLMGAAKGLQSLPHTDPSPLQRYSEDPV-PLPSETGYVAPLTCSPQE	1132
1027	GH-----PVRENSITURNISDPTQNALEKDLQGH-----E	1056
1133	YVYQPDVVRPOP-----PSPRE-----GPIIP-AARPAGATLERAKTSLPGKNGVWKDVF	1179
1057	YVYQPGSTSSRLSDIYNPNVEDLTDGWPVSLSSQEAETNFSRPEYLVNTQNSL---PL	1113
1180	AFGAVENPEYLVTPQGGAAPOPHPPPAFSPAFDNLVYWDQPPPERGAPPSTFKGTPTAEN	1239
1114	VSSGSMDDPDY---QAG-----YQAAF-----LPQTGALTGMGMLPAAEN	1151

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:22:34 ; Search time 48.3575 Seconds
(without alignments)
5347.444 Million cell updates/sec

Title: SEQ4-250-264-12
Perfect score: 6801
Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYGLDVPV 1255

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL 21.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rvirus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6203	91.2	1259	6 O18735	O18735 canis famil
2	3124	45.9	1208	11 Q90X70	Q90X70 ratus norv
3	3095	45.5	1210	11 Q9EP98	Q9EP98 mus musculu
4	2737	40.2	1165	13 Q9YH40	Q9YH40 xiphophorus
5	2696.5	39.6	1137	13 Q9W6F6	Q9W6F6 gallus gall
6	2287	33.6	1328	13 P79754	P79754 fugu rubrip
7	2002.5	29.4	1433	5 Q9BTH9	Q9BTH9 anopheles g
8	1782.5	26.2	419	4 Q9UK79	Q9UK79 homo sapien
9	1739	25.6	367	11 Q8R2X1	Q8R2X1 mus musculu
10	1720	25.3	729	15 Q8G712	Q8G712 avian rous-
11	1718	25.3	567	15 Q8G714	Q8G714 avian rous-
12	1697.5	25.0	412	4 Q8WV0	Q8WV0 homo sapien
13	1653.5	24.3	962	15 Q64895	Q64895 avian eryth
14	1645	24.2	545	15 Q85468	Q85468 avian eryth
15	1486.5	21.9	655	11 Q9WVF5	Q9WVF5 mus musculu
16	1470.5	21.6	643	11 Q9ERV6	Q9ERV6 mus musculu

17	1251	18.4	1193	5 Q9YIX8	Q9YIX8 ephydatia f
18	1167	17.2	1717	5 Q26566	Q26566 xhistosoma
19	1158.5	17.0	1368	5 Q23821	Q23821 caenorhabdi
20	1108	16.3	527	13 Q90836	Q90836 gallus gall
21	981.5	14.4	478	11 Q9ESE0	Q9ESE0 ratus norv
22	973.5	14.3	599	13 Q9PSH2	Q9PSH2 gallus gall
23	906	13.3	165	4 Q14256	Q14256 homo sapien
24	887	13.0	176	11 Q923V5	Q923V5 ratus norv
25	806.5	11.9	346	13 P11776	P11776 xiphophorus
26	778	11.4	435	5 Q8SZW1	Q8SZW1 drosophila
27	754.5	11.1	311	13 Q99162	Q99162 xiphophorus
28	752.5	11.1	1362	13 Q9PVZ4	Q9PVZ4 xenopus lae
29	729	10.7	1671	5 Q9NJV5	Q9NJV5 biophalar
30	702.5	10.3	1368	13 Q8UW85	Q8UW85 paralicth
31	694	10.2	331	4 Q9BUD7	Q9BUD7 homo sapien
32	687	10.1	1418	13 Q93457	Q93457 scophthalmu
33	684.5	10.1	1369	13 Q8UW86	Q8UW86 paralicth
34	669	9.8	1358	13 Q73798	Q73798 xenopus lae
35	652.5	9.6	1472	5 Q9U5A8	Q9U5A8 bombyx mori
36	648	9.5	1412	13 Q8UW84	Q8UW84 paralicth
37	646	9.5	1245	13 Q9YGH8	Q9YGH8 scophthalmu
38	636.5	9.4	1418	13 Q8UW83	Q8UW83 paralicth
39	631	9.3	149	6 Q9SG66	Q9SG66 oryctolagus
40	622	9.1	1371	11 Q9QVW4	Q9QVW4 ratus sp.
41	613.5	9.0	2144	5 Q9VD94	Q9VD94 drosophila
42	598	8.8	935	4 Q96L35	Q96L35 homo sapien
43	592	8.7	987	11 Q91YMO	Q91YMO mus musculu
44	589	8.7	987	11 Q99MR2	Q99MR2 mus musculu
45	587.5	8.6	1036	4 Q07912	Q07912 homo sapien

ALIGNMENTS

RESULT 1

O18735 ID O18735 PRELIMINARY; PRT; 1259 AA.
AC O18735;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Erbb-2.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RA Yokota H.;
RT "cDNA cloning of erbb-2 from canine mammary gland.";
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB008451; BAA23127.1; -
DR HSSP; P11362; 1FGK.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR InterPro; IPR004019; YLP_motif.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR Pfam; PF02757; YLP_2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS00018; EF_HAND; UNKNOWN 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; transferase; tyrosine-protein kinase.
SQ SEQUENCE 1259 AA; 137989 MW; E37364D49C4ACD46 CRC64;

```
Query Match 91.2%; Score 6203; DB 6; Length 1259;
Best Local Similarity 91.0%; Pred. No. 0;
Matches 1147; Conservative 41; Mismatches 66; Indels 6; Gaps 2;

QY 1 MELAALCRWGLLLALLPPGAASQVCTGTDMKRLPASPEHLDMLRHLVYGGQVVOGNNL 60
DB 1 MELAANCRWGLLLALLPPGAAGTQVCTGTDMKRLPASPEHLDMLRHLVYGGQVVOGNNL 60
QY 61 ELTYLPTNASLFLQDIQEVGVYVLIHNOVROVPLQRLIRVGTQLFEDNYALAVLDNG 120
DB 61 ELTYLPTNASLSLFLQDIQEVGVYVLIHNOVROVPLQRLIRVGTQLFEDNYALAVLDNG 120
QY 121 DPLNNTPTVTCASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILMKDIFHKNNOLA 180
DB 121 DPLNNTPTVTCASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILMKDIFHKNNOLA 180
QY 181 LTIIDNRSRACHPCSMKSGRCWGBSSBDCQSLTRTVCGGCARCKGPLPTDCCHQC 240
DB 181 LTIIDNRSRACHPCSMKSGRCWGBSSBDCQSLTRTVCGGCARCKGPLPTDCCHQC 240
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVOYIKANSKFIGITELRYTFGASCVTACP 300
DB 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTSCP 300
QY 301 YNYLSTDVSGCTLVCPILHNOEVAEDGTQRCCKSKPCARVCYCLGMEHLREVRVTSAN 360
DB 301 YNYLSTDVSGCTLVCPILHNOEVAEDGTQRCCKSKPCARVCYCLGMEHLREVRVTSAN 360
QY 361 IQPFAGCKKIFGSLAFIPESFDGDPASNTAPLOPEQLQVFETLEETIGYLIYSAPDPSLP 420
DB 361 IQPFAGCKKIFGSLAFIPESFDGDPASNTAPLOPEQLRVFEALEETIGYLIYSAPDPSLP 420
QY 421 DLSVFONLVIRGRILHNGAYSUTLOGLGSWLGRLSLRELGLALIHNNHLCFVHTV 480
DB 421 NLSVFONLVIRGRILHNGAYSUTLOGLGSWLGRLSLRELGLALIHNNHLCFVHTV 480
QY 481 PWQLFNPQHALLHTANRDECEVCGSLACHOLCARGHCWGPGPTQCVNCSOFLRGOEC 540
DB 481 PWQLFNPQHALLHSANREBECVGBGLACYP-CAHGHCHWGPGPTQCVNCSOFLRGOEC 539
QY 541 VESCRVLQGLPREYVNRHCLCPHCEQONGSVTCFGEADQVCAHYKDPFPFCVARC 600
DB 540 VESCRVLQGLPREYVNRHCLCPHCEQONGSVTCFGEADQVCAHYKDPFPFCVARC 599
QY 601 PSQVKPDLSPYMTWKPPDESGACQPCINCTHSDVLDKGCAPORASPLTSIVSAVVG 660
DB 600 PSQVKPDLSPYMTWKPPDESGACQPCINCTHSDVLDKGCAPORASPLTSIVSAVVG 659
QY 661 ILLVVLGVVFGILIKRROQKIRKYTMRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720
DB 660 ILLVVLGVVFGILIKRROQKIRKYTMRRLLQETELVEPLTPSGAMPNQAMRILKETEL 719
QY 721 RKVKVLGSAFGTYVGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVWAGVSP 780
DB 720 RKVKVLGSAFGTYVGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVWAGVSP 779
QY 781 YVSRLLGICLTSTVQLTQVLMYPCCLLDHVRNREGRGLSGDILLNWCQIAKGSYLEYDR 840
DB 780 YVSRLLGICLTSTVQLTQVLMYPCCLLDHVRNREGRGLSGDILLNWCQIAKGSYLEYDR 839
QY 841 LVHRDLAARNVLKSPNHVKITDIFGLARLLDIDETEHADGKVPKIMWALESIILRRFT 900
DB 840 LVHRDLAARNVLKSPNHVKITDIFGLARLLDIDETEHADGKVPKIMWALESIILRRFT 899
QY 901 HQSDVMSYGVTVVLMETFGAKPDGIPAREIPDLLEKGERLPPOPICTIDVYIMVVKCM 960
DB 900 HQSDVMSYGVTVVLMETFGAKPDGIPAREIPDLLEKGERLPPOPICTIDVYIMVVKCM 959
QY 961 IDSECRPRRELVAESRMARDPQRFVIONEDLGPASPLDSTFYRSLLEDDMDGLVDA 1020
DB 960 IDSECRPRRELVAESRMARDPQRFVIONEDLGPASPLDSTFYRSLLEDDMDGLVDA 1019
QY 1021, EYLVPQQGFFCPDPAPGAGMVRHRRSSSTRSGGGDLTLGLEPSEEPKSLAPSEG 1079
DB 1020 EYLVPQQGFFCPDPAPGAGMVRHRRSSSTRSGGGDLTLGLEPSEEPKSLAPSEG 1079
QY 1081 AGSDVDFDGLMCAAKGLSLPHTDPSPLQRYSEDTVPPLPSETDGYVAPLTCSPQPEYV 1140
DB 1080 AGSDVDFDGLMCAAKGLSLPHTDPSPLQRYSEDTVPPLPSETDGYVAPLTCSPQPEYV 1139
QY 1141 NOPDVRPQPPSPREGPLPAARPAATLER-----AKTLSPGKGVVVDVPAFGGAVENPE 1195
DB 1140 NQDEWVWPQPLALEGLPPSPRPAATLERPKTLSPKTLSPGKGVVVDVPAFGGAVENPE 1199
QY 1196 YLTPOGGAAPQPHPPPAFPADNLYWDDQPPERGAPSTFKGTATENPEYGLDVPV 1255
DB 1200 YLAPRGRAAPQPHPPPAFPADNLYWDDQPPERGAPSTFKGTATENPEYGLDVPV 1259

RESULT 2
Q90X70 PRELIMINARY; PRT; 1209 AA.
ID 090X70;
AC 090X70;
DT 01-MAY-2000 (T-EMBLrel. 13, Created)
DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
DE Epidermal growth factor receptor.
GN EGFR.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=101116;
RN [1]
RC SEQUENCE FROM N.A.
RX MEDLINE=90358888; PubMed=2342466;
RA Petch L.A.; Harris J.; Raymond V.W.; Blasband A.J.; Lee D.C.;
RT "A truncated, secreted form of the epidermal growth factor receptor is
RT encoded by an alternatively spliced transcript in normal rat tissue.";
RL Mol. Cell. Biol. 10:2973-2982(1990).
RN [2]
RC SEQUENCE FROM N.A.
RX STRAIN=FISHER; TISSUE=LIVER;
RA Petch L.A.;
RT Submitted (NOV-1991) to the EMBL/GenBank/DBJ databases.
RN [3]
RC SEQUENCE FROM N.A.
RX STRAIN=FISHER; TISSUE=LIVER;
RA Guttridge K.; Dawson T.L.; Earp H.S.;
RT Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RX HSP; P11362; 1FGK.
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; Pkinase; 1.
DR Pfam; PF01030; Recept_L domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TyKc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00111; PROTEIN KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; Receptor; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 1209 AA; 134891 MW; 96FE7F6C1B7773 CRC64;

Query Match 45.9%; Score 3124; DB 11; Length 1209;
Best Local Similarity 49.9%; Pred. No. 2.5e-226;
Matches 637; Conservative 166; Mismatches 360; Indels 114; Gaps 26;

QY 3 LAALCRWGLLLALLPPGA-ASTQVCTGTDMKRLPASPEHLDMLRHLVYGGQVVOGNNLE 61
DB 3 LAALCRWGLLLALLPPGA-ASTQVCTGTDMKRLPASPEHLDMLRHLVYGGQVVOGNNLE 61
```

Db	15	LAALCAAG-----GALBEKKVCQGTSSNRILTQGTPEDHFLSLQRMFNNECVVLGNLE	66
Qy	62	LTYLPTNASLSFLODIQEOVGXYLIAHNVQRQVPLQRLURIVRGTQLFPEDNYALAVLNDGD	121
Db	67	ITYVQRNYDL5FLAKTIQEVAGVGLIALNVTVERIPLENLIQIRGNALYENTYALAVLSN--	124
Qy	122	PLNNTTPTVGASPGGLRELQLRSLTEILKGGVILQORNPOLCYQDITILWKDIFHKNNQAL	181
Db	125	-----YGTNKTGLRELPRNLQEBILIGAVRFSNNPILCNMETIQRDIV-QDVFLSN	175
Qy	182	TLIDITNRS-RACHPCSPMKSGRSGWESSBDCQSLTRTVCAAGCA-RCKGPLPTDCCHEQ	239
Db	176	MSMDVQRHLTCPCPKDPSCPNGSGWGRGEENCQKLTIIIAQQCSRRCRGSPSDCCHNQ	235
Qy	240	CAAGCTGPKHSDCLACILFHNHSGICELHCPALVOYIKANSKFIGITELRYTFGASCVTAC	299
Db	236	CAAGCTGPRESDCLVCHRDEATCKDTCPLMLYNPTTYQMDVNPPEKYSFGATCVKKC	295
Qy	300	PYNYLSTDVSGCTLVCPLHNOEVTABDGTQRCCKSKPCARVCYGLGMEHLREVRATVSA	359
Db	296	PRNTVVTDHGSCVRACGPDYIEV--BEDGVSCKKGDGFCRVKNGIGIGEFKDTLSINAT	354
Qy	360	NIQFAGCKKIFGSLAFUPESGDGPASNTAPLOPEQLQVFETLEEITGYLVIISAMPDSL	419
Db	355	NIKFKYCTAISGDHLILPFAFKGDSFTRTPLOPRELEIUKTVKEITGFLIIQAWPEN4	414
Qy	420	PDL5VFQNLQVIRGILRHNGAYSLTLOGLIGSWLGRSLRBLGSLALIHHTHLCFVHT	479
Db	415	TDLHAFENLEIIRGTRKHQGFSLAVGLNITSILGRSLKEISDGDVIIISGRNLCYANT	474
Qy	480	VPWDLFRNPHOALLHTANRDEDCVGBGLACHOLCARGHCGPGPTQVCNCSQFLRQOE	539
Db	475	INWKLFQTPNQKTKIMNRAEKDCKATNHVCNPLCSSEGCGWGPETDCVSCQNVSRGRE	534
Qy	540	CVESRVLQGLPREVYNARHCLPCHPECOPQNGSVTCFGPADQCVACAHYKDPFCVAR	599
Db	535	CVDKNILGEGPREFVENSECIIQCHPECLPOTMNIITCTGRGPDNCIIKAAHYVDGPHCVKT	594
Qy	600	CPSGVKPDL5YMPYTKWFPDEBEGAQPCPCINCTHSCVDLDDKGCDAEQRASPLTSIVSAV	658
Db	595	CPSGIMGENNTL-VKMFADANNVCHLCHANCTYGCAGPLKGC--QQPEGPKISITATGI	651
Qy	659	VGILLVVVLGVVFGI-LIKRPOQKIRKYTMRLLOETELVPLTPSGAMPNOAQWRLIKE	717
Db	652	VGGILLFIW-VALGTGLFMRRRQLVRKRTLRLQLQERELVEPLTPSGEAPNOAHLRIKE	710
Qy	718	TELKVKVLGSGAGCTVYKGTWIDGENVKIPVAIKVLRENTSPKANKEILDEAYMVAGV	777
Db	711	TEFKKIKVLGSGAGTVYKGLWIIPEGEKVIIPVAIKELREATSPKANKEILDEAYMVASV	770
Qy	778	GSPVYSRLIGICLSTVOLTQMLPYGCLLDHVRNRRGLSGQDLNWCMOIAKMSYLE	837
Db	771	DNPHVCRLLGICLSTVQLITQMLPYGCLLDYVREHKDNISGOYLLNWCVOIAKGMVYLE	830
Qy	838	DVRLVHRDLAARNVLKSPNHVKITDIFGLARLLDIDETEHADGCKVPKIKWMALESILRR	897
Db	831	DRRLVHRDLAARNVLKTPQVKITDIFGLAKLLGAEKEYHAEGKVKPIKWMALESILHR	890
Qy	898	RPTHQSDVMSYGVVTWELMTGKAPYDGI PAREIPDLLEKGERLPPOPPICTIDVYIMVVK	957
Db	891	IYTHQSDVMSYGVVTWELMTGSKPYDGIIPASEISSILEKGERLPPOPPICTIDVYIMVVK	950
Qy	958	CWMTDSECRPRELVSFEFSMARDPQRFFVIO--NEDLGPASPLDSTPYRSLLEDODMDG	1016
Db	951	CWMTDAGSRPKFREILIEFSKMAKDPQRYLVIQDGERMHLPSPTDSNFPYALMBEEDMD	1010
Qy	1017	LVDAAEYLVPOQGFPCPDPAFGWGVHHRHSSTRSGGGDLTLGLEPSEEEAPRSPLA	1076
Db	1011	VVDADYELVPOQGF-----NSPST-----SRPPLL	1036
Qy	1077	PSEGAGSDVFDGDLGMGAAGLQSLPHDPSPLOQRYSDPTVPVLPSET--DGYVAPITCS	1134
Db	1037	SSLSANEN-----SSTVACINRNGSCRKVEDAFLOQYSDPTSVLETDNIDDTFL-----	1086

Qy	1135	PQBYVNPQDVRCQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVVKDVFATCGAVENP	1139
Db	1087	PVPEYINQ-SVPRKPRAGSVQNYHYNQPLHP-----APGRDLHYQN--PHSNAVSNP	1135
Qy	1195	EYL-TPCGGAAPQHPPHAFSPAFDNLNLYWDQ-----DP-----PERGAPSTF	1237
Db	1136	EYLNIAQ-----PTCLSSGFDSSALMIQKSHQMSLDNPDYQDDFFPKAKPNGIF	1186
Qy	1238	KGTTPTAENPEYGLDVP 1254	
Db	1187	KG-PTAENAEYLRVAPP 1202	
RESULT 3			
Q9EP98			
ID	Q9EP98	PRELIMINARY; PRT; 1210 AA.	
AC	Q9EP98		
DT	01-MAR-2001 (TrEMBLrel. 16, Created)		
DT	01-MAR-2001 (TrEMBLrel. 16, Last sequence update)		
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)		
DE	Epidermal growth factor receptor isoform 1.		
GN	EGFR.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
NCBI_TaxID=10090;			
UN	[1]		
SEQUENCE FROM N.A.			
STRAIN=C3H/101, 129/SVJ, AND 129/SVEVITAC;			
Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,			
Sinclair C.S., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,			
Balasubramanian S., Crossley T.O., Magnuson T.R., James C.D.,			
Maihle N.J.;			
RT	"Comparative genomic sequence analysis and isolation of human and		
RT	mouse alternative Egr transcripts encoding truncated receptor		
RT	isoforms."		
RL	Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.		
UN	[2]		
SEQUENCE FROM N.A.			
STRAIN=C57BL/6J;			
Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,			
Schehl C., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,			
Balasubramanian S., Crossley T.O., Magnuson T.R., James C.D.,			
Maihle N.J.;			
RT	"Comparative genomic sequence analysis and isolation of human and		
RT	mouse alternative Egr transcripts encoding truncated receptor		
RT	isoforms."		
RL	Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.		
UN	[3]		
DR	EMBL; AF275366; AAC28045.1; -		
DR	EMBL; AF275364; AAC28045.1; JOINED.		
DR	EMBL; AF275365; AAC28045.1; JOINED.		
DR	EMBL; AF275367; AAC24386.1; -		
DR	HSSP; P11362; 1FGK.		
DR	MGD; MGI:95294; Egfr.		
DR	InterPro; IPR000345; Cyt c heme bind.		
DR	InterPro; IPR000494; EGFR_L_domain.		
DR	InterPro; IPR000719; Euk_pkinase.		
DR	InterPro; IPR002174; Furin-like.		
DR	InterPro; IPR002290; Ser thr_pkinase.		
DR	InterPro; IPR001245; Tyr_pkinase.		
DR	Pfam; PF00757; Furin-like; 1.		
DR	Pfam; PF00059; pkinase; 1.		
DR	Pfam; PF01030; Recep_L_domain; 2.		
DR	PRINTS; PR00109; TYRKINASE.		
DR	ProDom; PD000001; Euk_pkinase; 1.		
DR	SMART; SM00261; FU; 5.		
DR	SMART; SM00220; S_TKc; 1.		
DR	SMART; SM00219; TyKc; 1.		
DR	PROSITE; PS00190; CYTOCHROME C; UNKNOWN 1.		
DR	PROSITE; PS00107; PROTEIN KINASE ATP; 1.		
DR	PROSITE; PSS0011; PROTEIN KINASE DOM; 1.		
DR	PROSITE; PS00109; PROTEIN KINASE TYR; 1.		


```
Qy 549 GLPREY-VNARHCLPCHPEQOPONGSVTCFGEADOCVACAHYKDPFCVVARCP----- 601
Db 507 SLRPLYSVDSKTCGDCHQCKD-----FCYGPNEGNCSCMNVDGRFCFVACPTTKHAM 561
Qy 602 -----SGVKPDLSPYMTWPKFPD----- 618
Db 562 NGTCINCHKTCVGRGRPRDTIAPDGCISCDAKIIGSDAKIERCLMKDSCPDGYSDYVL 621
Qy 619 -EG----- 621
Db 622 QEEGPLQLSGKAVCRKCHPRCKKCTGYGFHEQFCQECTGYKXGEQCECPQDFYANEE 681
Qy 622 --ACQCPINCT-----HSCVDL-----DD-----KGCPEAQ----- 646
Db 682 TRICLPCHQCRGCHGLGDHHECRNLKLFEGDPYDNATFTCVSNCPASHPKYRPOEA 741
Qy 647 -----RASPLTSIVSAVVGILVVLGVVFGI---LKRROQKIRKYTM 687
Db 742 GKIGPYCSADSMQSLRIEPTQVKIVMGSGMALILLCVVFGIAFLFSRHKNKDAVKM 801
Qy 688 RRLQETELVEPLTPSGAMPNQMRILKETELRKVKVLGSGAGFTYKGIWIPDGENVK 747
Db 802 TMLAGCEDSELRPSNVGNLTKRIKEAIRRGVGLWGAFGRVFKGVMPGESVK 861
Qy 748 IPVAIKVIRENTSPKANKEILDAYVMAGVSPVSRLLIGLCTSTVQLVTQMLPYGCLL 807
Db 862 IPVAIKVLEMSGSESEKFELEAYINASVEHNLKLLAVCTSQMWLITQLMPJGCLL 921
Qy 808 DHVRENRGLSGODLNCWQIAKGMYSLEDRVLVHRDLAARNVLVSPNHVKITPFGLA 867
Db 922 DYVRNNKDKIGSKALLNWSQIARGMAYLEERLVRDLAARNVLVQTPSCVKITVFGLA 981
Qy 868 RLIDIDETEYHAGGKPIKMALESILRRRFTHQSDVWVGVTWELMTFGAKPYDGP 927
Db 982 KLLDFDSDEYRAGGKMPIKWLALCEIRHVF7TSKSDVWAFGTTIMELLTYGARPVENP 1041
Qy 928 AREIPDLLEGERLPQPICTIDVYIMVWCWIMIDSECRPRFRFELVSEFSRMARDPORFV 987
Db 1042 AKDVELIEIGHKLPPQDICSLDVYCIILSCWLDADARTFQLAETFAEKARDPGRYL 1101
Qy 988 VIQNEGLGASPLDSTPYRSLLEDDMGDLV----- 1018
Db 1102 MI-----PGDKFMRLPSYTNQDEKDLIRTLAPVMAAAAAAAGASNVDPSTIA 1152
Qy 1019 DAREYLPQOGFCPCDPAGCAGGVHHRSSSTRSGGDLTLCLEPSEBAPRS----- 1073
Db 1153 ETEYLOPKTRPSIMLPESA-----GVGGIR-----VEPS-DEMPKSLRYCK 1188
Qy 1074 -PLAP---SEGAGSDVPDGLGMAKGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVA 1129
Db 1189 DPLKPDDETGHGKEV-----GVGGIR-----LNLPLDEDDYLM 1222
Qy 1130 PLTCSPOPEYVNPQDVRPQPPSPREGPLPAARPAGATLERAKTLPSPKNGVWVDVAFG 1189
Db 1223 P-TCQSQ---NQS-----TPG-----YMDLIGVPA 1243
Qy 1190 AVENPEYL-----TPQGAAPQHPBPAPSPAFDNLVYWDQDPPERGAPPSIFKGT 1240
Db 1244 SVNPEYLMGSGTQAIAGLAQSGMG--DHTPP-----PNTPNMG 1280
Qy 1241 PTAENPE 1247
Db 1281 PTHQHSQ 1287

RESULT 8
Q9UK79 ID Q9UK79 PRELIMINARY; PRT; 419 AA.
AC Q9UK79;
DT 02-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Herstatin.
```

```
GN HER-2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99415951; PubMed=10485918;
RA Doherty J.K., Bond C., Jardim A., Adelman J.P., Clinton G.M.;
RT "The HER-2/neu receptor tyrosine kinase gene encodes a secreted
RT autoinhibitor.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:10869-10874 (1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Doherty J.K., Clinton G.M., Adelman J.P., Evans A.J., Henner W.D.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF177761; AAD56009.2; -.
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR002174; Furin-Like.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L domain; 1.
DR SMART; SM00261; FU; 1.
SQ SEQUENCE 419 AA; 45472 MW; FEC1BE347E2D030C CRC64;

Query Match 26.2%; Score 1782.5; DB 4; Length 419;
Best Local Similarity 86.3%; Pred. No. 6.6e-126;
Matches 340; Conservative 7; Mismatches 40; Indels 7; Gaps 2;

Qy 1 MELAALCRWGLLLALLPPGAASTQVCTGTMKRLRLPASPTHLDMLRHLYQGCVVQGNL 60
Db 1 MELAALCRWGLLLALLPPGAASTQVCTGTMKRLRLPASPTHLDMLRHLYQGCVVQGNL 60
Qy 61 EUTYLPNTASLSFLODIQSVQGVLIHAHQVQPLQRLRVIRGTOLFEDNYALAVLDNG 120
Db 61 EUTYLPNTASLSFLODIQSVQGVLIHAHQVQPLQRLRVIRGTOLFEDNYALAVLDNG 120
Qy 121 DPLNNTPTVTGASPGGLRELQLRLSLTEILKGGVLIQORNQOLCYQDTILWKDIFHKNNQLA 180
Db 121 DPLNNTPTVTGASPGGLRELQLRLSLTEILKGGVLIQORNQOLCYQDTILWKDIFHKNNQLA 180
Qy 181 LTLIDNTRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVCAAGCARCKGPLPTDCHEQC 240
Db 181 LTLIDNTRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVCAAGCARCKGPLPTDCHEQC 240
Qy 241 AGCTGPKHSDCLACLHNHSGICELHCPALVOYIKANSKFITELRYTFGASCVTACP 300
Db 241 AGCTGPKHSDCLACLHNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Qy 301 YNYLSTDVGSCTILVCPPLHNQEVTAEDGTORCEKSKPCARVCYGLGMEHLREVRAVTSAN 360
Db 301 YNYLSTDVGSCTILVCPPLHNQEVTAEDGTORCEKSKPCAR----GTHSLPPRPAVPVP 355
Qy 361 IOEFAGCKKIFGSLAPLPESFDGDPASNTAPLOP 394
Db 356 LRMQPG--PAHPVLSFLRPSDLVSAFYSLPLAP 387

RESULT 9
Q9R2X1 ID Q9R2X1 PRELIMINARY; PRT; 367 AA.
AC Q9R2X1;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Hypothetical 40.2 kDa protein.
OS Mus musculus (Mouse);
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
```



```
DR ENBL; BC027080; AAH27080.1; -.
KW Hypothetical protein.
SQ SEQUENCE 367 AA; 40163 MW; 0BE03395F9E101B0 CRC64;

Query Match      25.6%; Score 1739; DB 11; Length 367;
Best Local Similarity 88.0%; Pred. No. 1e-122;
Matches 32; Conservative 11; Mismatches 33; Indels 0; Gaps 0;

Qy 889 MALESILRRRFTHQSDVMSYGVTVWELMTFGAKYDGIPIAREIPDLLEKGERLPQPPICT 948
Db 1 MALESILRRRFTHQSDVMSYGVTVWELMTFGAKYDGIPIAREIPDLLEKGERLPQPPICT 60

Qy 949 IDVTMIMVKCMIDSECRPRRELVSFSEMRARDPQRFVIVQNEIDLGPASPLDSTFVRSL 1008
Db 61 IDVTMIMVKCMIDSECRPRRELVSFSEMRARDPQRFVIVQNEIDLGPSPMDSTFVRSL 120

Qy 1009 LEDDMDGLVDABEYLVPOQGFCDPAPGAGGMVHRHRSRSPRSRGGDLTLGLEPSEE 1068
Db 121 LEDDMGELVDABEYLVPOQGFSPDPAFGTGTSTHRRHRSRSPRSRGGDLTLGLEPSEE 180

Qy 1069 EAPRSLPASEGAGSDVFDGLGMAAKGLQSLFTHDPSPLQRYSEDPTVPLPSETDGYV 1128
Db 181 EPPRSLPASEGAGSDVFDGLAVGVTGKGLQSLSPHDLSPLORYSEDPTLPLPETDGYV 240

Qy 1129 APLTCSQPEYVNOQDVVRPPSPREGPLPAAPAGATLERAKTLSPGKGVVVDVPAFG 1188
Db 241 APLACSPQPEYVNOQDVVRPPSPREGPLPPEGPPPIRPAGATLERPKTLSPGKGVVVDVPAFG 300

Qy 1189 GAVENPEYLTPOGGAAPQHPHPPAFSPAFONLYWQDPPERGAPPTFGTPTAENPEY 1248
Db 301 GAVENPEYLAAPRAGTASQHPHPPAFSPAFONLYWQDQNSSEQGPPPTFGTPTAENPEY 360

Qy 1249 LGLDVPV 1255
Db 361 LGLDVPV 367

RESULT 10
Q86712 ID Q86712 PRELIMINARY; PRT; 729 AA.
AC Q86712;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Polypeptide.
GN POLYPROTEIN.
OS Avian rous-associated virus type 1.
OC Viruses; Retroviral viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11950;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94203659; PubMed=8152791;
RA Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
RA Johnson A., Beug H.;
RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
RT evolution of distinct viral genomes carrying mutant v-erbB genes with
RT different transforming capacities.";
RL Oncogene 9:1307-1320(1994).
DR ENBL; S69372; AAC60725.1; -.
DR HSP; P03322; IAKS.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR004028; Retro M.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00011; PROTEIN KINASE DOM; 1.
DR PROSITE; PS00109; PROTEIN KINASE_TYR; 1.
KW ATP-binding, Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 729 AA; 80649 MW; 84D2F6914EFE1D63 CRC64;
```

```
Query Match      25.3%; Score 1720; DB 15; Length 729;
Best Local Similarity 54.8%; Pred. No. 7.7e-121;
Matches 358; Conservative 75; Mismatches 118; Indels 102; Gaps 15;

Qy 569 PQNGSVTCGPEADQCVACAHYKDPFCVACRPSGVKPDLSYMPIKFPDDEGACQCPPI 628
Db 141 PEETATPKTGP--DHCMKCAHFIDGPHCVKACPAVLGENDTL-VWKYADANAVCQLCHP 197

Qy 629 NCTHSCVDLDDKCPAEQASPLTISVAVV-GILLVWVVLGVVFGILIKRRQOKIRKYTM 687
Db 198 NCTRGCKGPLEGCP--NGSKTPSTAAGVVGGLLCLVVVGLGIGLYLRR-HIVKRTL 253

Qy 688 RLLQTELVEPLTPPGAMPNQAMRILKETELRKVKVLGSGAFGVYKGIWIPOGENVK 747
Db 254 RLLQRELVEPLTPSGEAPNQAHRLILKETETPKVKVLGSGAFGVYKGLWIPGEKVK 313

Qy 748 IPVAIKVLRENTSPKANKETLDEAYVMAGVGVSPVSRLLGICLTSTVOLVTLMPYGCCL 807
Db 314 IPVAIKELREATSPKANKETLDEAYVMASVDNPRVRLGICLTSTVOLVTLMPYGCCL 373

Qy 808 DHVRENRLGSLQDLNLCWQIAKMSYLEVDRLVHRDLAARNVLKSPNHVKITDFGLA 867
Db 374 DYIREHKDNLGSOYLLNWCQIAKGNVLEERLVRDLAARNVLKTPQHVKITDFGLA 433

Qy 868 RLLDIDETEHADGGKVPKIKWMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIP 927
Db 434 KLLGADEKEYHAEGGKVPKIKWMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIP 493

Qy 928 AREIPDLLEKGERLPQPPICTIDVTMIMVKCMIDSECRPRRELVSFSEMRARDPQRFV 987
Db 494 ASEISSVLKGERLPQPPICTIDVTMIMVKCMIDADSRPKRELIAEFSKMARDPQRYL 553

Qy 988 VIO-NEDLGPASPLDSTFYRSILLEDDMDGLVDABEYLVPOQGFCDPAPGAGGMVHR 1046
Db 554 VIQDERMHLPSPTDSKFTYRLMEEDMEDIVDAEYLVPHQGF-----598

Qy 1047 HRSSSTRSGGDLTLGLEPSEEEAPRSL-----APSEGAGSDVFDGLGMAAKGLQSL 1101
Db 599 -NSPST-----SRTPLLSLSATSNNSATNCID-----RNGQGH 631

Qy 1102 PTHDPSPLQRYSEDPTVPLPSET--DGYVAPLTCSPQPEYVNOQDVVRPPSPREGPLPA 1159
Db 632 PVREDSFVQRYSSDPTGTFLEESIDGDL-----PAPEYVNO--LMPKKPS-----675

Qy 1160 ARPAGATLERAKTLSPGKGVVVDV-----AFGGAVENPEYL 1197
Db 676 -----TAMVQNIYNNISLTAISKLPMDSRYNQNSHSTAVDNPYL 715

RESULT 11
Q86714 ID Q86714 PRELIMINARY; PRT; 567 AA.
AC Q86714;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE V-erbB protein (Fragment).
GN V-ERBB.
OS Avian rous-associated virus type 1.
OC Viruses; Retroviral viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11950;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94203659; PubMed=8152791;
RA Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
RA Johnson A., Beug H.;
RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
RT evolution of distinct viral genomes carrying mutant v-erbB genes with
RT different transforming capacities.";
RL Oncogene 9:1307-1320(1994).
DR ENBL; S69372; AAC60727.1; -.
DR HSP; P11362; IFGK.
DR InterPro; IPR000719; Euk_pkinase.
```

```
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00069; kinase.1.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_kinase; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS0011; PROTEIN KINASE DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW Tyrosine-protein kinase.
FT NON_TER 1
SQ SEQUENCE 567 AA; 63390 MW; C6D9CBA7ADF725B1 CRC64;

Query Match 25.3%; Score 1718; DB 15; Length 567;
Best Local Similarity 55.4%; Pred. No. 7.5e-121;
Matches 357; Conservative 73; Mismatches 112; Indels 102; Gaps 15;

QY 578 GPSADQVCAHYKDPFCVARGSPGVKPLSYMPIWKFDPDEGACQPCINCHSCVDL 637
DB 1 GP--DHCMKCAHFDGPHCVKACPAVLGENDTL-VWKYADANAVCOLCHPNCTRGCKGP 57
QY 638 DDKGCPAEQRASPLTSTVSADV-GILLVVLGVVFGILIKRRQKIRKTYMRLLQETEL 696
DB 58 GLGECPP---NGSKTPSIAAGVGGLLCLVVGIGLYLRRR-HIVKRTLRLLQEREL 113
QY 697 VEPLTPSGAMPNOAQMILKETELRKVKVLGSGAFGVYKGIWIPDGENVKIPIVAIKVLR 756
DB 114 VEPLTPSGEAPNQAHLRIKETEFKVKVVLGSGAFGVYKGLWIPGEKVKIPIVAIKELR 173
QY 757 ENTSPKANKILDEAYVMAGVSGPYVSRLLIGICLTSTVQLVLTQMPYGCILLDHVRNRR 816
DB 174 EATSPKANKILDEAYVMASVDNPRVCRLLIGICLTSTVQLITQMPYGCILLDVIKHKON 233
QY 817 LGSODLLNWCQITAKGMSYLEDLVRLDIAARNVLKSPNHVKITDPLGLARLLDDETE 876
DB 234 IGSQYLLNWCQITAKGMYLEERLVRDLAARNVLKTPQHKITDPLGLAKLLGADEKE 293
QY 877 YHADGGKVPKIMMALESILRRRTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLE 936
DB 294 YHAEGGKVPKIMMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIPIAREIPDLLE 936
QY 937 KGERLPOPPCTIDVYIMVKCWMIDSECRPRFRELVSERFARMARDPQRFVVIQ-NEDLG 995
DB 354 KGERLPOPPCTIDVYIMVKCWMIDADSRPKFRELIAEFSKWARDPPRYLVITQGBRMH 413
QY 996 PASPLDSTFYRSLLEDDMDGLVDAEYLVPOQGFCCPDAPGAGGMVHRRHSSSTRSG 1055
DB 414 LPSTDSKFRYLWEEEDMDIVDAEYLVPHOGFF-----NSPST--- 454
QY 1056 GGDLTGLPESEEAAPRSPL-----APSEGAGSDVDFDGLMGAAKGLQSLPHDPSPLQ 1110
DB 455 -----SRTPLLSLSATSNNATNCID-----RNGQGHVPVREDSFVQ 491
QY 1111 RYSEDFTVPLPSET--DGYVAPITCSQPYVYNQPDVRPQPPSPREGPLPAARPAGATLE 1168
DB 492 RYSSDPTGNFLESIDGFL-----PAPEYVQ--LMPKKPS----- 526
QY 1169 RAKTLPFGKNGVKDVF-----AFGGAVENPEYL 1197
DB 527 -----TAMVQOYNNISLTAKLPMDSRYQNSHSTAVDNPEYL 566

RESULT 12
Q8WYV0
ID Q8WYV0 PRELIMINARY; PRT; 412 AA.
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 44.7 kDa protein.
GN P36559.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Primates; Catarrhini; Hominidae; Homo.

InterPro; IPR001245; Tyr_kinase.
Pfam; PF00069; kinase.1.
PRINTS; PR00109; TYRKINASE.
ProDom; PD000001; Euk_kinase; 1.
SMART; SM00219; TyrKc; 1.
PROSITE; PS00107; PROTEIN KINASE ATP; 1.
PROSITE; PS0011; PROTEIN KINASE DOM; 1.
PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
Tyrosine-protein kinase.
NON_TER 1
SEQUENCE 567 AA; 63390 MW; C6D9CBA7ADF725B1 CRC64;

Query Match 25.3%; Score 1718; DB 15; Length 567;
Best Local Similarity 55.4%; Pred. No. 7.5e-121;
Matches 357; Conservative 73; Mismatches 112; Indels 102; Gaps 15;

QY 578 GPSADQVCAHYKDPFCVARGSPGVKPLSYMPIWKFDPDEGACQPCINCHSCVDL 637
DB 1 GP--DHCMKCAHFDGPHCVKACPAVLGENDTL-VWKYADANAVCOLCHPNCTRGCKGP 57
QY 638 DDKGCPAEQRASPLTSTVSADV-GILLVVLGVVFGILIKRRQKIRKTYMRLLQETEL 696
DB 58 GLGECPP---NGSKTPSIAAGVGGLLCLVVGIGLYLRRR-HIVKRTLRLLQEREL 113
QY 697 VEPLTPSGAMPNOAQMILKETELRKVKVLGSGAFGVYKGIWIPDGENVKIPIVAIKVLR 756
DB 114 VEPLTPSGEAPNQAHLRIKETEFKVKVVLGSGAFGVYKGLWIPGEKVKIPIVAIKELR 173
QY 757 ENTSPKANKILDEAYVMAGVSGPYVSRLLIGICLTSTVQLVLTQMPYGCILLDHVRNRR 816
DB 174 EATSPKANKILDEAYVMASVDNPRVCRLLIGICLTSTVQLITQMPYGCILLDVIKHKON 233
QY 817 LGSODLLNWCQITAKGMSYLEDLVRLDIAARNVLKSPNHVKITDPLGLARLLDDETE 876
DB 234 IGSQYLLNWCQITAKGMYLEERLVRDLAARNVLKTPQHKITDPLGLAKLLGADEKE 293
QY 877 YHADGGKVPKIMMALESILRRRTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLE 936
DB 294 YHAEGGKVPKIMMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIPIAREIPDLLE 936
QY 937 KGERLPOPPCTIDVYIMVKCWMIDSECRPRFRELVSERFARMARDPQRFVVIQ-NEDLG 995
DB 354 KGERLPOPPCTIDVYIMVKCWMIDADSRPKFRELIAEFSKWARDPPRYLVITQGBRMH 413
QY 996 PASPLDSTFYRSLLEDDMDGLVDAEYLVPOQGFCCPDAPGAGGMVHRRHSSSTRSG 1055
DB 414 LPSTDSKFRYLWEEEDMDIVDAEYLVPHOGFF-----NSPST--- 454
QY 1056 GGDLTGLPESEEAAPRSPL-----APSEGAGSDVDFDGLMGAAKGLQSLPHDPSPLQ 1110
DB 455 -----SRTPLLSLSATSNNATNCID-----RNGQGHVPVREDSFVQ 491
QY 1111 RYSEDFTVPLPSET--DGYVAPITCSQPYVYNQPDVRPQPPSPREGPLPAARPAGATLE 1168
DB 492 RYSSDPTGNFLESIDGFL-----PAPEYVQ--LMPKKPS----- 526
QY 1169 RAKTLPFGKNGVKDVF-----AFGGAVENPEYL 1197
DB 527 -----TAMVQOYNNISLTAKLPMDSRYQNSHSTAVDNPEYL 566

RESULT 13
Q64895
ID Q64895 PRELIMINARY; PRT; 962 AA.
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Gag v-erb-A, v-erb-B protein.
DE GAG, v-erb-A, v-erb-B.
GN Avian erythroblastosis virus.
OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=11861;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90206603; PubMed=1969616;
RA Bruskin A., Jackson J., Bishop J.M., McCarley D.J., Schatzman R.C.;
RT "Six amino acids from the retroviral gene gag greatly enhance the transforming potential of the oncogene v-erb-B.";
RT
```

Onco gene 5:15-24(1990).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR EMBL; X52209; CAA36459.1; -
DR EMBL; X52211; CAA36459.1; JOINED.
DR HSP; P10828; 2NLL.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR000536; Hormone_rec_lig.
DR InterPro; IPR001723; Steroid_receptor.
DR InterPro; IPR001245; Tyr_kinase.
DR InterPro; IPR001628; Znf_C4steroid.
DR Pfam; PF00104; hormone_rec; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00105; zf-C4; 1.
DR PRINTS; PR00398; STRDHORMONER.
DR PRINTS; PR00047; STROIDFINGER.
DR ProDom; PD000001; Euk_pkinase; 1.
DR ProDom; PD000035; Znf_C4steroid; 1.
DR SMART; SM00430; HOL1; 1.
DR SMART; SM00219; TyrKc; 1.
DR SMART; SM00399; Znf_C4; 1.
DR PROSITE; PS00031; NUCLEAR RECEPTOR; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00011; PROTEIN KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN KINASE_TYR; 1.
KW ATP-binding; DNA-binding; Nuclear protein; Receptor;
KW Transcription regulation; Nuclease; Tyrosine-protein kinase;
KW Zinc-finger.
SQ SEQUENCE 962 AA; 108320 MW; 3C5AED791E4E95CE CRC64;

Query Match 24.3%; Score 1653.5; DB 15; Length 962;
Best Local Similarity 51.7%; Pred. No. 1.2e-115;
Matches 358; Conservative 73; Mismatches 142; Indels 119; Gaps 18;

QY 541 VEECRVLQGLPRE-YVNAH-HCLP-----CHPEQC 568
DB 354 IEKQESYLAFEHYINRKHNIPIHFWSKLIMKVADLRMTGAYHASFHMKVCEPTELS 413
QY 569 PQNGSVTCFGEADQCACAHYKDPFCVACPSGVKPDLSYMPIWKFPEDEGACQPCPI 628
DB 414 PQE-----VGP--DHCWKCAHFIDGPHCVKACPAVLGENDTL-VWKYADANAVCQLCHP 465
QY 629 NCTHSCVDLDDKGPAPQASPLTSIVSAVV-GILLVVVLGVVFGILIKRQKIRKYTM 687
DB 466 NCTRGCKGPGLEGCP---NGSKTPSIAAGVVGGLCLVVGIGLGLYLR--HIVRKRTL 521
QY 688 RLLQETELVEPLTPSGAMPNQAMRILKETELRKVKVLGSGAGTGVYKGIWIPDGENVK 747
DB 522 RLLQETELVEPLTPSGAMPNQAMRILKETELRKVKVLGSGAGTGVYKGIWIPDGENVK 581
QY 748 IPVAIKVLRNTSPKANKEILDEAYVMAGVSPVYVRLGLCLTSTVQLTQLMFYGCLL 807
DB 582 IPVAIKVLRNTSPKANKEILDEAYVMAGVSPVYVRLGLCLTSTVQLTQLMFYGCLL 641
QY 808 DHVRENRGLGSDLLNWCQIAKMSYLEDLVRLVHRLAARNVLKSPNHNKIPFGLA 867
DB 642 DYIREHNDKNGSQVLLNWCQIAKMSYLEDLVRLVHRLAARNVLKSPNHNKIPFGLA 701
QY 868 RLIDIDETEYHAGGKVPKIKMALESILRRFTHQSDVMSYGVTVWELMTFGAKPDGIP 927
DB 702 KQLGADEKEYHAGGKVPKIKMALESILRRFTHQSDVMSYGVTVWELMTFGAKPDGIP 761
QY 928 ARBIPDLLEKGERLPQPICTIDVYIMVWKCMIDSECRPFRELSEFSEFMRDPRFV 987
DB 762 ASISSVLEKGERLPQPICTIDVYIMVWKCMIDSECRPFRELSEFSEFMRDPRFV 821
QY 988 VIQ-NEDLGASPLDSTFYRSLLEDDMDGLVDAEYLVPOQGFPCPDAPGAGVMVHR 1046
DB 822 VIQGDERMHLPSPTDSKFKYRTLMEEDMEDIVDAEYLVPHQGF-----866
QY 1047 HRSSSTRSGGDLTLGLPESEBAPRSLAPSEGAGSDVDFDGLGMAKGLQSLTHDP 1106
DB 867 -NSPST-----SRTPLLSSLSATSN-----NSATKCIDRNGH--898

QY 1107 SPLQRYSEDPTVPLPSETGTYVAPLTCSPQPEYVNDQVVRPQPPSPREGPLPAARPAAT 1166
DB 899 -----PVREDGFL-----PAPEYVNO--LMPKKPSTAMVQNYINVISLT 936
QY 1167 -LERAKTLPKNGVGVKDFAFGGAVENTPEYL 1197
DB 937 AISKLPMDSRYNQ-----SHSTAVDNPEYL 961

RESULT 14
Q85468 ID Q85468 PRELIMINARY; PRT; 545 AA.
AC Q85468;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Avian Erythroblastosis virus (Ts34) v-erbB gene.
OS Avian erythroblastosis virus.
OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=11861;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88217326; PubMed=2897102;
RA Scotting P., Vennstrom B., Jansen M., Graf T., Beug H., Haymann M.J.;
RT "Common site of mutation in the erbB gene of avian erythroblastosis
RT virus mutants that are temperature sensitive for transformation.";
RL Oncogene Res. 1:265-278(1987).
DR EMBL; X06943; CAA30024.1; -
DR HSP; P11362; IFGK.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00109; PROTEIN KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN KINASE_TYR; 1.
KW ATP-binding; Tyrosine-protein kinase.
SQ SEQUENCE 545 AA; 60899 MW; 140DCB8CCA0F8AF4 CRC64;

Query Match 24.2%; Score 1645; DB 15; Length 545;
Best Local Similarity 54.9%; Pred. No. 2.2e-115;
Matches 345; Conservative 70; Mismatches 121; Indels 92; Gaps 15;

QY 578 GREADOCVACAHYKDPFCVACPSGVKPDLSYMPIWKFPEDEGACQPCPINCTHSCVDL 637
DB 1 GP--DHCWKCAHFIDGPHCVKACPAVLGENDTL-VWKYADANAVCQLCHPNCCTRGCKP 57
QY 638 DDKGCPAEQASPLTSIVSAVV-GILLVVVLGVVFGILIKRQKIRKYTMRLLOETEL 696
DB 58 GLEGCP---NGSKTPSIAAGVVGGLCLVVGIGLGLYLR--HIVRKTLRLLQEREL 113
QY 697 VBLPTPSGAMPNQAMRILKETELRKVKVLGSGAGTGVYKGIWIPDGENVKIPVAIKVLR 756
DB 114 VBLPTPSGAMPNQAMRILKETELRKVKVLGSGAGTGVYKGIWIPDGENVKIPVAIKVLR 173
QY 757 ENTSPKANKEILDEAYVMAGVSPVYVRLGLCLTSTVQLTQLMFYGCLLDHVRNCR 816
DB 174 EATSPKANKEILDEAYVMAGVSPVYVRLGLCLTSTVQLTQLMFYGCLLDHVRNCR 233
QY 817 LGSQDILLNWCQIAKMSYLEDLVRLVHRLAARNVLKSPNHNKIPFGLA 876
DB 234 IGSQDILLNWCQIAKMSYLEDLVRLVHRLAARNVLKSPNHNKIPFGLA 293
QY 877 YHAGGKVPKIKMALESILRRFTHQSDVMSYGVTVWELMTFGAKPDGIPAREIPDLLE 936
DB 294 YHAGGKVPKIKMALESILRRFTHQSDVMSYGVTVWELMTFGAKPDGIPAREIPDLLE 353
QY 937 KGERLPQPICTIDVYIMVWKCMIDSECRPFRELSEFSEFMRDPRFV 995
DB 354 KGERLPQPICTIDVYIMVWKCMIDSECRPFRELSEFSEFMRDPRFV 413

QY	996	PASPLDSTFYRSLLEDDMGDLVDABEYLVPQGGFFCPDPAAGAGGVMHRRHSSSTRSG	105
Db	414	LPSPDTSKKFYRTLMEEDMEDIVADAEYLVPHQGGF-----NSPST---	454
QY	1056	GGDLTLGLEPSEEEAPRSL-----APSEGAGSDVFDGDLGMAAGLQSLPHDPSPLO	1110
Db	455	-----SRTPLLSLSATSNNSATNCIDRNGG-----H-----	481
QY	1111	RYSEDPVPLPSETGDGVVAPLTCSPQPEYVYNQPDVRPQPPSPREGPLPAARPAAGAT-LER	1169
Db	482	-----PVREDGFL-----PAPEYVYNQ-LMPEKPESTAMVQIQIYNISLTAISK	523
QY	1170	AKTILSPGKGVGVKDVAFAGGAVENPEYL	1197
Db	524	LPMDSRYN-----SHSTAVDNPEYL	544

RESULT 15

Q9WVF5

ID

Q9WVF5

PRELIMINARY;

PRT;

655 AA.

AC

Q9WVF5;

DT

01-NOV-1999

(T-EMBLrel. 12, Created)

DT

01-NOV-1999

(T-EMBLrel. 12, Last sequence update)

DT

01-JUN-2002

(T-EMBLrel. 21, Last annotation update)

DE

Epidermal growth factor receptor

(Epidermal growth factor receptor isoform 3).

DE

EGFR.

OS

Mus musculus (Mouse).

OC

Eukaryota; Metazoa;

OC

Mammalia; Eutheria; Rodentia;

OC

NCBI_TaxID=10090;

RN

[1]

RP

SEQUENCE FROM N.A.

RC

STRATN=C57BL/6J; TISSUE=LIVER;

RA

Reiter J.L.; Threadgill D.W.; Danielsen A.J.; Schell C.; Lampland A.L.; Balasubramaniam S.; Crossley T.O.; Magnuson T.R.; Maibhe N.J.;

RA

"Alternative Transcripts from the Human and Mouse EGFR Genes Encode Carboxy-Terminal Truncated Receptors.";

RL

Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.

RN

[2]

RP

SEQUENCE FROM N.A.

RC

STRATN=C3H/101, 129/SVJ, AND 129/SVEVTCAC;

RA

Reiter J.L.; Threadgill D.W.; Eley G.D.; Strunk K.E.; Danielsen A.J.; Sinclair C.S.; Pearsall R.S.; Green P.J.; Yee D.; Lampland A.L.; Balasubramaniam S.; Crossley T.O.; Magnuson T.R.; James C.D.; Maibhe N.J.;

RA

"Comparative genomic sequence analysis and isolation of human and mouse alternative Egr transcripts encoding truncated receptor isoforms.";

RL

Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.

RN

[3]

RP

SEQUENCE FROM N.A.

RC

STRATN=C57BL/6J; TISSUE=LIVER;

RX

MEDLINE=21085660; Pubmed=11217851;

RA

Kawai J., Shingagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y., Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S., Azawa K., Ozaki Y., Nishi K., Kiyosawa H., Kondo S., Yamanaka I., Saito T., Itakura M., Nishi K., Gojobori T., Bono H., Kasukawa T., Saito R., Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T., Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H., Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J., Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T., Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G., Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F., Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M., Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H., Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P., Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N., Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F., Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L., Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S., Hayaashi-aki Y.;

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:05:49 ; Search time 63.1875 Seconds
(without alignments)
1909.378 Million cell updates/sec

Title: US-09-806-703A-4
Perfect score: 6812
Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYGLDVPV 1255

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 73:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6806	99.9	1255	1 A24571	protein-tyrosine k
2	5988	87.9	1260	1 TVRTNU	protein-tyrosine k
3	5984.5	87.9	1254	2 I48161	p-185 precursor -
4	3168	46.5	1210	1 GQHUE	epidermal growth f
5	3144	46.2	1210	2 A53183	epidermal growth f
6	3123.5	45.9	1223	1 TVCHLV	epidermal growth f
7	3003.5	44.1	1308	2 A47253	epidermal growth f
8	2701	39.7	1166	1 S06142	protein-tyrosine k
9	2431.5	35.7	1342	2 A36223	kinase-related tra
10	2346.5	34.4	1339	2 JC4387	epidermal growth f
11	1766.5	25.9	698	1 TVFVLV	protein-tyrosine k
12	1703	25.0	604	1 TVYHVV	protein-tyrosine k
13	1652.5	24.3	1330	1 GQFPE	epidermal growth f
14	1647	24.2	544	2 S35745	protein-tyrosine k
15	1640	24.1	545	2 S00727	kinase-related tra
16	1623	23.8	540	2 B44776	protein-tyrosine k
17	1621	23.8	540	1 TVFVEB	protein-tyrosine k
18	1536	22.5	644	2 A36325	epidermal growth f
19	1302	19.1	1323	2 B88257	protein let-23 (im
20	1302	19.1	1374	2 S70712	protein-tyrosine k
21	1214	17.8	1369	2 S70713	protein-tyrosine k
22	1177	17.3	1717	1 A45558	epidermal growth f
23	1155	17.0	527	2 A42032	epidermal growth f
24	997.5	14.6	843	2 A27131	epidermal growth f
25	806.5	11.8	346	2 S13807	protein-tyrosine k
26	754.5	11.1	311	2 T43220	insulin-like growt
27	735	10.8	1363	2 T43220	insulin receptor p
28	718	10.5	1382	1 INHUR	insulin receptor p
29	711	10.4	1383	2 A36080	insulin receptor p

RESULT 1

A24571

protein-tyrosine kinase (EC 2.7.1.112) erbB2 precursor - human

N;Alternate names: c-erb-B-2 protein precursor; kinase-related transforming protein erbB.

C;Species: Homo sapiens (man)

C;Date: 25-Oct-1987 #sequence revision 06-Dec-1996 #text change 11-Jun-1999

C;Accession: A24571; A25491; A4188; B4188; I59509; I57622

R;Yamamoto, T.; Ikawa, S.; Akiyama, T.; Semba, K.; Nomura, N.; Miyajima, N.; Saico, T.; J

Nature 319, 230-234, 1986

A;Title: Similarity of protein encoded by the human c-erb-B-2 gene to epidermal growth f

A;Reference number: A24571; MUID:86118663; PMID:3003577

A;Accession: A24571

A;Molecule type: mRNA

A;Residues: 1-1255 <YAM>

A;Cross-references: GB:X03363; NID:G31197; PIDN:CAA27060.1; PID:G31198

R;Semba, K.; Kanata, N.; Toyoshima, K.; Yamamoto, T.

Proc. Natl. Acad. Sci. U.S.A. 82, 6497-6501, 1985

A;Title: A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB-1/epiderm

A;Reference number: A25491; MUID:86016729; PMID:2995967

A;Accession: A25491

A;Molecule type: DNA

A;Residues: 737-1031 <SEM>

A;Cross-references: GB:M11767; NID:G182163; PIDN:AAA35808.1; PID:G553282

R;Cousseus, L.; Yang-Feng, T.L.; Liao, Y.C.; Chen, E.; Gray, A.; McGrath, J.; Seeburg, P

Science 230, 1132-1139, 1985

A;Title: Tyrosine kinase receptor with extensive homology to EGF receptor shares chromos

A;Reference number: A44188; MUID:86070181; PMID:2999974

A;Accession: A44188

A;Molecule type: DNA

A;Residues: 740-910 <COU1>

A;Cross-references: GB:M12036; NID:G183988; PIDN:AAA35978.1; PID:G183989

A;Accession: B44188

A;Molecule type: mRNA

A;Residues: 1-517, 'RALL', 522, 'S', 524-654, 'V', 656-1169, 'A', 1171-1255 <COU2>

A;Cross-references: GB:M11730; NID:G183986

R;King, C.R.; Kraus, M.H.; Aaronson, S.A.

Science 229, 974-976, 1985

A;Title: Amplification of a novel v-erbB-related gene in a human mammary carcinoma.

A;Reference number: I59509; MUID:85272597; PMID:2992089

A;Accession: I59509

A;Status: translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 832-909 <REX>

A;Cross-references: GB:I29395; NID:G459807; PIDN:AAA35809.1; PID:G459808

R;Tal, M.; King, C.R.; Kraus, M.H.; Ullrich, A.; Schlessinger, J.; Givol, D.

Mol. Cell. Biol. 7, 2597-2601, 1987

A;Title: Human HER2 (neu) promoter: evidence for multiple mechanisms for transcriptional

A;Reference number: I57622; MUID:87286898; PMID:3039351

A;Accession: I57622

A;Status: translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-191 <TAL>

A;Cross-references: GB:M16792; NID:g183983; PIDN:AAA58637.1; PID:g553332
 C;Comment: Amplification and overexpression of this erbB-related gene occurs in about 30
 C;Genetics:
 A;Gene: GDB:ERBB2; NCL; NEU; HER-2
 A;Cross-references: GDB:120613; OMIM:164870
 A;Map position: 17q21.1-17q21.1
 A;Introns: 25/1; 75/3; 147/1; 983/3
 A;Note: the list of introns is incomplete
 C;Function:
 A;Description: catalyzes the phosphorylation of a peptidyl tyrosine residue by ATP
 C;Superfamily: epidermal growth factor receptor; protein kinase homology
 C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
 inase
 F;1-21/Domain: signal sequence #status predicted <SIG>
 F;22-1255/Product: protein-tyrosine kinase erbB2 #status predicted <MAT>
 F;22-653/Domain: extracellular #status predicted <EXT>
 F;70-304/Domain: EGF receptor extracellular domain repeat <EE1>
 F;395-605/Domain: EGF receptor extracellular domain repeat <EE2>
 F;654-675/Domain: transmembrane #status predicted <TM>
 F;718-983/Domain: intracellular #status predicted <INT>
 F;726-734/Region: protein kinase homology <KIN>
 F;68/124,187,259,530,571,629/Binding site: carbohydrate (Asn) (covalent) #status predict
 F;686/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
 F;753/Active site: Lys #status predicted
 F;1139,1221,1222,1248/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation)

Query Match 99.98; Score 6806; DB 1; Length 1255;
 Best Local Similarity 99.88; Pred. No. 1.2e-270;
 Matches 1253; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY	1	MELAAALCRWGLLALLP	GAAS	QVCTG	TDKMLR	PAS	PTH	LDMLR	HLH	YQG	QV	QGNL	60
DB	1	MELAAALCRWGLLALLP	GAAS	QVCTG	TDKMLR	PAS	PTH	LDMLR	HLH	YQG	QV	QGNL	60
QY	61	ELTYLPTN	ASLS	FLQ	IQE	VG	YVLI	IAH	NO	VR	QV	PLR	120
DB	61	ELTYLPTN	ASLS	FLQ	IQE	VG	YVLI	IAH	NO	VR	QV	PLR	120
QY	121	DPLNNTPT	VTG	AS	PG	GLR	EL	QL	SL	RL	EL	KG	180
DB	121	DPLNNTPT	VTG	AS	PG	GLR	EL	QL	SL	RL	EL	KG	180
QY	181	LT	LID	NR	S	R	A	C	P	S	C	M	240
DB	181	LT	LID	NR	S	R	A	C	P	S	C	M	240
QY	241	A	A	C	T	G	P	K	H	S	D	C	300
DB	241	A	A	C	T	G	P	K	H	S	D	C	300
QY	301	Y	N	L	S	T	D	V	G	S	C	T	360
DB	301	Y	N	L	S	T	D	V	G	S	C	T	360
QY	361	I	Q	E	F	A	G	C	K	I	F	G	420
DB	361	I	Q	E	F	A	G	C	K	I	F	G	420
QY	421	D	L	S	V	F	Q	N	L	Q	V	I	480
DB	421	D	L	S	V	F	Q	N	L	Q	V	I	480
QY	481	P	W	D	L	R	N	P	H	O	A	L	540
DB	481	P	W	D	L	R	N	P	H	O	A	L	540
QY	541	V	E	S	C	R	V	L	Q	G	L	P	600
DB	541	V	E	S	C	R	V	L	Q	G	L	P	600
QY	601	P	S	G	V	K	P	D	L	S	M	T	660
DB	601	P	S	G	V	K	P	D	L	S	M	T	660

RESULT 2

TVRTNU

protein-tyrosine kinase (EC 2.7.1.112) neu precursor - rat

C;Species: Rattus norvegicus (Norway rat)

C;Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 11-Jun-1999

C;Accession: A24562; A61204

R;Bargmann, C.I.; Hung, M.C.; Weinberg, R.A.

Nature 319, 226-230, 1986

A;Title: The neu oncogene encodes an epidermal growth factor receptor-related protein.

A;Reference number: A24562; MUID:86118662; PMID:3945311

A;Accession: A24562

A;Molecule type: mRNA

A;Residues: 1-1260 <BAR>

A;Cross-references: EMBL:X03362; NID:g56745; PIDN:CAA27059.1; PID:g56746

R;Masui, T.; Mann, A.M.; Macatee, T.L.; Garland, E.M.; Okamura, T.; Smith, R.A.; Cohen, S.

Carcinogenesis 12, 1975-1978, 1991

A;Title: Direct DNA sequencing of the rat neu oncogene transmembrane domain reveals no m

2-thiazolyl]formamide or N-methyl-N-nitrosourea.

A;Reference number: A61204; MUID:92035293; PMID:1682063

A;Accession: A61204

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 637-663, 'V', 665-702 <MAS>

A;Note: authors translated the codon GCA for residue 25 as Val

C;Genetics:

A;Gene: neu

C;Superfamily: epidermal growth factor receptor; protein kinase homology

C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho

F;1-19/Domain: signal sequence #status predicted <SIG>

F;20-1260/Product: protein-tyrosine kinase neu #status predicted <MAT>

F:658-680V/Domain: transmembrane #status predicted <TMN>
F:723-988/Domain: protein kinase homology <KIN>
F:731-739/Region: protein kinase ATP-binding motif
F:711,263,535,576,634/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:691/Binding site: phosphate (Thr) (covalent) #status predicted
F:758/Active site: Lys #status predicted
F:882,1227,1253/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 87.9%; Score 5988; DB 1; Length 1260;
Best Local Similarity 87.7%; Pred. No. 2.8e-237;
Matches 1103; Conservative 50; Mismatches 102; Indels 2; Gaps 2;

QY 1 MELAALCRWGLLALLPPGAASQVCTGDMKRLPASPETHLDMLRHLYQGCVVQGNL 60
DB 4 MELAALCRWGLLALLPPGAGTAGVCTGDMKRLPASPETHLDMLRHLYQGCVVQGNL 63
QY 61 ELTYLPTNASLSFLQDIQEVQGYVLIHANOVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
DB 64 ELTYVVPANASLSFLQDIQEVQGYVLIHANOVRQVPLQRLRIVRGTQLFEDNYALAVLDNR 123
QY 121 DPLNNTPTVT-GASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDTILWKDIFHKNNQL 179
DB 124 DPQDNVAASPTGRTEGLELQRLSLTEILKGGVLIQORNPOLCYQDMVLWKDFRKNQNL 183
QY 180 ALTLIDTNRSRACHPCSPMCKGRWCSESDQSLTRTVACGACRCKGPLETDCHEQ 239
DB 184 APVDIDTNRSRACHPCSPMCKGRWCSESDQSLTRTVACGACRCKGPLETDCHEQ 243
QY 240 CAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTAC 299
DB 244 CAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTTC 303
QY 300 PYNLSTDVSGSCTLVCPPLHNOEYTAEDGTQCEKSKPCARVCYGLGMEHLRVRVTS 359
DB 304 PYNLSTEVSGSCTLVCPPLHNOEYTAEDGTQCEKSKPCARVCYGLGMEHLRVRVTS 363
QY 360 NIOEFAGCKIKFSLAPLPSFGDPASNTAPLOEQVFTLEITGLYTSAMPDLSL 419
DB 364 NVQEFDFGCKIKFSLAPLPSFGDPASNTAPLOEQVFTLEITGLYTSAMPDLSL 423
QY 420 PDLVSFONLQVIRGRLIHNCAYSLTLQGLIGISWLGRLSRLGSLALIHNNHLCFVHT 479
DB 424 RDLVSFONLQVIRGRLIHNCAYSLTLQGLIGISWLGRLSRLGSLALIHNNHLCFVHT 483
QY 480 VPWDQLFRNPHQALLTANRPEDE-CVGEGLACHQLCARGHCGPPTQCVNCSQFLRGQ 538
DB 484 VPWDQLFRNPHQALLTANRPEDE-CVGEGLACHQLCARGHCGPPTQCVNCSHFLRGQ 543
QY 539 ECVEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFQPEADQCVACAHYKDPDFCVA 598
DB 544 ECVEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFQPEADQCVACAHYKDPDFCVA 603
QY 599 RCPGSGVKPDLVSYPIMKFPDEEGACOPCPINCTHSCVDLDDKCPAEORASPLTISVAV 658
DB 604 RCPGSGVKPDLVSYPIMKFPDEEGACOPCPINCTHSCVDLDDKCPAEORASPLTISVAV 663
QY 659 VGILLVVLVGVGGLIKRQKIRKYTWRRLLQETELVEPLTPSGAMPNQAOIRILKET 718
DB 664 EGVLFLVVLVGVGGLIKRQKIRKYTWRRLLQETELVEPLTPSGAMPNQAOIRILKET 723
QY 719 ELRKVKVLGSGAGTGVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMAGVG 778
DB 724 ELRKVKVLGSGAGTGVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMAGVG 783
QY 779 SPVVSRLGICLTSTVOLVTLQMPYGLLDHVENRGLSGDQLLKNWCQIAKMSYLE 838
DB 784 SPVVSRLGICLTSTVOLVTLQMPYGLLDHVENRGLSGDQLLKNWCQIAKMSYLE 843
QY 839 VRLVHRDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGCKVPKIMMALESILRR 898
DB 844 VRLVHRDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGCKVPKIMMALESILRR 903
QY 899 FTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYMINVKC 958

DB 904 FTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYMINVKC 963
QY 959 WMIDSECRFRRELVSERFARMARDPQRFVVIQNEIDLGASPLDSTFYRSLLEDGMDLV 1018
DB 964 WMIDSECRFRRELVSERFARMARDPQRFVVIQNEIDLGASPLDSTFYRSLLEDGMDLV 1023
QY 1019 DAEYLVPOQGFPCPDPAFGAGVMHRRSSSTSGGDLTLGLSEPEEAPRSPAPS 1078
DB 1024 DAEYLVPOQGFPCPDPAFGAGVMHRRSSSTSGGDLTLGLSEPEEAPRSPAPS 1083
QY 1079 EGAGSDVDFGDLGMAKGLQSLPHTDPSPIQRYSEDVPLPSETDGVVAPLTCSPQE 1138
DB 1084 EGAGSDVDFGDLGMAKGLQSLPHTDPSPIQRYSEDVPLPSETDGVVAPLTCSPQE 1143
QY 1139 YNQSDVDFGDLGMAKGLQSLPHTDPSPIQRYSEDVPLPSETDGVVAPLTCSPQE 1198
DB 1144 YNQSDVDFGDLGMAKGLQSLPHTDPSPIQRYSEDVPLPSETDGVVAPLTCSPQE 1203
QY 1199 POGGAAPHPHPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGIDVFP 1255
DB 1204 POGGAAPHPHPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGIDVFP 1260

RESULT 3
148161
p-185 precursor - golden hamster
C:Species: Mesocricetus auratus (golden hamster)
C>Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 18-Jun-1999
C:Accession: I48161
R:Nakamura, T.; Ushijima, T.; Ishizaka, Y.; Nagao, M.; Arai, M.; Yamazaki, Y.; Ishikawa, Gene 140, 251-255, 1994
A:Title: Cloning and activation of the Syrian hamster neu proto-oncogene.
A:Reference number: I48161; MUID:94193007; PMID:7908275
A:Accession: I48161
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1254 <RES>
A:Cross-references: GB:D16295; NID:G493236; PIDN:BAA03801.1; PID:G747595
C:Genetics:
A:Gene: neu
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP
F:718-983/Domain: protein kinase homology <KIN>
F:726-734/Region: protein kinase ATP-binding motif

Query Match 87.9%; Score 5984.5; DB 2; Length 1254;
Best Local Similarity 87.6%; Pred. No. 3.8e-237;
Matches 1099; Conservative 58; Mismatches 97; Indels 1; Gaps 1;

QY 1 MELAALCRWGLLALLPPGAASQVCTGDMKRLPASPETHLDMLRHLYQGCVVQGNL 60
DB 1 MELAALCRWGLLALLPPGAASQVCTGDMKRLPASPETHLDMLRHLYQGCVVQGNL 60
QY 61 ELTYLPTNASLSFLQDIQEVQGYVLIHANOVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120
DB 61 ELTYLPTNASLSFLQDIQEVQGYVLIHANOVRQVPLQRLRIVRGTQLFEDNYALAVLDNR 120
QY 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDTILWKDIFHKNNQL 180
DB 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDTILWKDIFHKNNQL 180
QY 181 LTLIDTNRSRACHPCSPMCKGRWCSESDQSLTRTVACGACRCKGPLETDCHEQ 240
DB 181 LTLIDTNRSRACHPCSPMCKGRWCSESDQSLTRTVACGACRCKGPLETDCHEQ 240
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTAC 300
DB 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTTC 300
QY 301 PYNLSTDVSGSCTLVCPPLHNOEYTAEDGTQCEKSKPCARVCYGLGMEHLRVRVTS 360
DB 301 PYNLSTDVSGSCTLVCPPLHNOEYTAEDGTQCEKSKPCARVCYGLGMEHLRVRVTS 360

A;Note: the EGF receptor (and other tyrosine kinases) can nick double-stranded DNA
 Cell 59, 33-43, 1989
 R;Chen, W.S.; Lazar, C.S.; Lund, K.A.; Welsh, J.B.; Chang, C.P.; Walton, G.M.; Der, C.J.
 A;Title: Functional independence of the epidermal growth factor receptor from a domain x
 A;Reference number: A33311; MUID:90003233; PMID:2790960
 A;Contents: annotation: internalization signal
 C;Comment: Binding of EGF to the receptor leads to internalization of the EGF-receptor c
 C;Genetics:
 A;Gene: GDB:EGFR
 A;Cross-references: GDB:120610; OMIM:131550
 A;Map position: 7p12.3-7p12.1
 C;Superfamily: epidermal growth factor receptor; protein kinase homology
 C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
 F;1-24/Domain: signal sequence #status predicted <SIG>
 F;25-1210/Product: EGF receptor #status predicted <MAT>
 F;25-645/Domain: extracellular #status predicted <EXT>
 F;75-300/Domain: EGF receptor extracellular domain repeat <E1>
 F;390-600/Domain: EGF receptor extracellular domain repeat <E2>
 F;646-668/Domain: transmembrane #status predicted <TM>
 F;669-1210/Domain: intracellular #status predicted <INT>
 F;710-726/Domain: protein kinase homology <KIN>
 F;718-726/Region: protein kinase ATP-binding motif
 F;999-1046/Region: coated-pit mediated internalization signal
 F;1047-1210/Region: inhibitory
 F;128,175,352,413,444,528,603/Binding site: carbohydrate (Asn) (covalent) #status predic
 F;745/Active site: Lys #status experimental

Query Match 46.5%; Score 3168; DB 1; Length 1210;
 Best Local Similarity 49.8%; Pred. No. 3e-122;
 Matches 630; Conservative 178; Mismatches 351; Indels 106; Gaps 21;

QY 11 LLALLPPGAA--STQVCTGDMKRLPASPETHDMLRLHYGCGQVQGNLETLVPTN 68
 DB 14 LLAALCPASALEBKVCQCTSNKLTQGTGFEDFLSLQRMFNCEVVLGNLEITYVRN 73
 QY 69 ASLSFLQIDIEVOGYVLIANQVQPLQRLRVRGTLQFEDNVALVLDNGPLNNTTP 128
 DB 74 YDLSFLKTIQEVAGYVLIANTVERIPLENLQIRGNMYENSVALAVLSNYD----- 126
 QY 129 VTGASPGGLRELQRLSLTEILKGVLIQFQNPOLCYQDITLWKDIFHKNQNALTLTIDNR 188
 DB 127 --ANKTGLKELPMRLNLOEILHGAVFNSNPALCNVESIQWRDILVSSDFLSNMSDFQNH 193
 QY 189 SRACHPCSPCKSGRCWGESSEDCQSLTRTVCAAGCA-RCKGPLPTDCHEQCAAGCTGP 247
 DB 184 LGSCQKCDPSCPCNGSCWGAEEENCQKLTIIQAQCCSGRCRGSFSDCCHNQCAAGCTGP 243
 QY 248 KHSDDLACLHFNHSGICELHCPALVTYNTDTFESMNPGRYTFGASCVTACPNYILSTD 307
 DB 244 RESDCLVCRKFRDEATCKDTCPLMLYNPTTYQMDVNPFGKYSFGATCVKKCPRYVYTD 303
 QY 308 VGSCTLVCPHLNQEVTAEDGTORCEKCKSPCARVCYGLGMEHLREVRVAVTSANIOEFAG 367
 DB 304 HGSQVACAGADSTEM--EEDGVKCKCKGCPKRCVNGIGIFGEFKDSUSINATIKHFKNC 362
 QY 368 KKIFGLSLAFSPESFDGDPASNTAPLQPEQLQVFTLEEITGYLISAWPDSLPDLSPFQN 427
 DB 363 TSISGDLHLPLVAFRGDSFHTPLDPLQELDLTKVKEITGFLLIQAWPENRTDLHAFEN 422
 QY 428 LQVIRGILNGAYSLTLOGLSISWLSRLSLRGSLALIHNTLHCFVHTVPWQOLFR 487
 DB 423 LEIRGTRKQHGQFSLAVWSLNTISLGRSLKEISDGDVITISGNKNCIYANTINWKKLFG 482
 QY 488 NPHQALLHTANRDEDECVGGLACHOLCARGHCGPPTOCVNCQSPQLRGQCEVEECRVL 547
 DB 483 TSQKTKIISNRGENSKATGQVCHALCSPGCGWPEPCVCRNVSRGRCVDCRCKLL 542
 QY 548 QGLPREYVNAHRLCPHCEQCPQNGSVTCFPEADQCAQAHYKDPFPFCVAPCSGVKPD 607
 DB 543 EGEPRFVENSEICQHCPECLPQAMNITCTGRPDNCIQCAHYIDGPHCVKTCPCAGWGE 602
 QY 608 LSYNPIWKFDPBAGACQPCPINCTHSCVDLDDKCPAEPASPLTSTISVAVG---ILLV 664

603 NNTL-VWKVADAGHVCHLCPNCTYCTGPGLEGCGTNGPKIP--SIATGMVGAALLLLV 659
 QY 665 VVLGVVFGILIKRRQKIRKYMRRLLOETELVEPLTPSGAMPNQAMRILKETELRKVK 724
 DB 660 VALGIG--LFMRHHIVKRTLRLLQRELVEPLTPSGEAPNQALLRIKETEFKKIK 716
 QY 725 VLGSAGFYVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDRAYVMAGVSGPVSR 784
 DB 717 VLGSAGFYVYKGLWIPGEKVKIPVAIKELREATSPKANKEILDRAYVMASVDNPHVC 776
 QY 785 LLGICLTSTVOLVTLQMPYGCILLDHVRENRLGSDLLNWCQIAKGSYLEDVRLVHR 844
 DB 777 LLGICLTSTVQLITQMPFGCLLDYVREKDNIGSYLLNWCQIAKGNWYLEDRLVHR 836
 QY 845 DLAAARNVLKSPNHVKITDFGLARLLDIDETEHADGKGVPIKMMMALESILRRRFTQSD 904
 DB 837 DLAAARNVLKTPQHVKITDFGLAKLLGAEKEYHABGGKVPKMMMALESILHRIYTHQSD 896
 QY 905 VMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPPICITIDVYIMVKCWMIDSE 964
 DB 897 VMSYGVTVWELMTFGSKPYDGIPIAREIPDLLEKGERLPQPPICITIDVYIMVKCWMIDAD 956
 QY 965 CRPREELVSESRWARDPQRFVVIQ--NEDLGPASPLDSTFVRSLLDDDDMGDLVDAAEY 1023
 DB 957 SRPKFRELIIETSKWARDPQRLVLIQGDRLMPLSPDTSNFRALMDEEDMDVDVDAEY 1016
 QY 1024 LVPQOGFFCPDPAPGAGGMVHRRHRSSTRSGGDLTLGLEPSEERAPRSLAPSEGAGS 1083
 DB 1017 LIPQOGFF-----SSPSTRTPLLSLSATS 1042
 QY 1084 DVPDGLGMAAKGLOSLTHDPSPLOQRYSEDTVPLPSET--DGVVAPLTCSPQPEYVN 1141
 DB 1043 N--NSTVACIDRNLQSCPKEDSFLOQYSSDPTGALTEDSIDDTL-----PVPEYIN 1094
 QY 1142 QDVRPQPSREGPLPAAPAGATLAKTLSPGKGVVQVFAFGGAVENPEYL-TPQ 1200
 DB 1095 Q--SVPKRPAQSVQNPVYHQNPLNP-----APSRDPHYQD--PHSTAVGNPEYLNVTQ 1143
 QY 1201 GGAAPQHPHPAFSPAFDNLVYWDQ-----DP-----PERGAPPSTFKGPTAE 1244
 DB 1144 -----PTCVNSTFDSPAWAKGSHQISLDNPDYQDDFPFKEAKPNKIFKGS-TAE 1193
 QY 1245 NPEYL 1249
 DB 1194 NAEYL 1198

RESULT 5
 A53183
 epidermal growth factor receptor precursor - mouse
 C;Species: Mus musculus (house mouse)
 C;Date: 06-Jan-1995 #sequence revision 06-Jan-1995 #text change 18-Jun-1999
 C;Accession: A53183; A43818; S24942; A28941; S45325; I49643
 R;Luetteke, N.C.; Phillips, H.K.; Qiu, T.H.; Copeland, N.G.; Earp, H.S.; Jenkins, N.A.; I
 Genes Dev. 8, 399-413, 1994
 A;Title: The mouse waved-2 phenotype results from a point mutation in the EGF receptor ty
 A;Reference number: A53183; MUID:94170986; PMID:8125255
 A;Accession: A53183
 A;Molecule type: mRNA
 A;Residues: 1-1210 <LUE>
 A;Cross-references: GB:U03425
 R;Avivi, A.; Lax, I.; Ullrich, A.; Schlessinger, J.; Givol, D.; Morse, B.
 Oncogene 6, 673-676, 1991
 A;Title: Comparison of EGF receptor sequences as a guide to study the ligand binding sit
 A;Reference number: A43818; MUID:91232866; PMID:2030916
 A;Accession: A43818
 A;Molecule type: mRNA
 A;Residues: 1-714 <AVI>
 R;Cross-references: GB:X59698
 R;Sisinger, D.P.; Serrero, G.
 submitted to the EMBL Data Library, June 1992
 A;Reference number: S24942
 A;Accession: S24942

A:Molecule type: mRNA
A:Residues: 969-971, 'K', 973-1115, 'D' <EIS>
A:Cross-references: EMBL:Z12608
R:Heisermann, G.J.; Gill, G.N.
J. Biol. Chem. 263, 13152-13158, 1988
A:Title: Epidermal growth factor receptor threonine and serine residues phosphorylated
A:Reference number: A28941; MUID:88330814; PMID:3138233
A:Accession: A28941
A:Molecule type: protein
A:Residues: 689-694, 'X', 696-704, 'L', 706-707, 989-992, 'XX', 995-996, 'X', 998-1000, 1002-1009,
R:Hibbs, M.L.; Dunn, A.R.; Alexander, W.S.
submitted to the EMBL Data Library, April 1994
A:Description: The complete cDNA sequence of the Mouse Epidermal Growth Factor Receptor
A:Reference number: S45325
A:Accession: S45325
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-971, 'K', 973-1210 <VER>
A:Cross-references: EMBL:X78987; NID:G488830; PIDN:CAA55587.1; PID:G488831
R:Paria, B.C.; Das, S.K.; Andrews, G.K.; Dev, S.K.
Proc. Natl. Acad. Sci. U.S.A. 90, 55-59, 1993
A:Title: Expression of the epidermal growth factor receptor gene is regulated in mouse b
A:Reference number: I49643; MUID:93126380; PMID:7678348
A:Accession: I49643
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 12-20, 122-132 <RES>
A:Cross-references: GB:L06864; NID:G193001; PIDN:AAA53029.1; PID:G567201
C:Genetics:
A:Gene: EGFR
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; growth factor receptor; kinase-related transforming protein; phosphop
F:1-24/Domain: signal sequence #status predicted <SIG>
F:648-670/Domain: transmembrane #status predicted <TM>
F:712-977/Domain: protein kinase homology <KIN>
F:720-728/Region: protein kinase ATP-binding motif
F:680,695/Binding site: phosphate (Thr) (covalent) #status experimental
F:697,1070,1071/Binding site: phosphate (Ser) (covalent) #status experimental
F:993/Binding site: (or 997) phosphate (Ser) (covalent) #status experimental
F:1028/Binding site: (or 1030 or 1032) phosphate (Ser) (covalent) #status experimental
F:1197/Binding site: phosphate (Tyr) (covalent) #status experimental

Query Match 46.2%; Score 3144; DB 2; Length 1210;
Best Local Similarity 49.8%; Pred. No. 2.8e-121;
Matches 633; Conservative 170; Mismatches 359; Indels 110; Gaps 23;
Qy 11 LLIALLPGAA--STQVCTGTDKMLRLPASPETHLMLRLHYLQCCQVQGNLELYLPTN 68
Db 14 LLTALCAAGGALEKKVCQGTSLRLTQGTGFEDHFLSLQRMVNNCEVVLGNLEITYVQRN 73
Qy 69 ASLSFLQDIOEVQGVYLIHNOVQVPLRLRIVRGTQLFEDNVALAVLQNGDPLNNTTP 128
Db 74 YDLSFLTKIOEVAGVYLIANTVERIPLENLQIIRGNALYENTVALAILSN----- 124
Qy 129 VTGASPGGLRELQLRLSLTEILKGGVLIQRPOLCYQDTILWKDI----FHKNNQLALTLI 184
Db 125 -YGNRTGRLRLPRLNLEILIGAVRFNNPILCNMDTIQWRDIVQNVFNNSMDL--- 180
Qy 185 DTVRSRACHPCSPMKSCRWGESSEDCQLTRTVAGGCA-RCKGPLPTDCHEQCAAG 243
Db 181 -QSPSSCPKCDPSCPNGSCWGGGEENCQKLTIKIACQCCSHRCGRSPSDCCCHNCAAG 239
Qy 244 CTGPKHSDCLACLFHNSGICELHCPALVTYNTDFTSPNPEGRYTFGASCVTACPNY 303
Db 240 CTGPRSDCLVCQFQDEATCKTCTPPLMYNPTYQMDVNPGEKYSFGATCKVKCPNY 299
Qy 304 LSTDVSGCTLVCPILHNOQVTAEDQTCRCKSPCARVCYGLGMEHLREVRVTSANIQE 363
Db 300 VVTDHGSCVRACGPDYVEV-EEDGIRKCKCDGCRKVCNGVIGIGEPKDTLSINATIKH 358
Qy 364 FAGCKKIFGSLAFPEFSDGPDASNTAPLOEQLOVFETLEETGYIYISAWPDSLPLDS 423
Db 359 FKYCTAISGDLHLPLFAFGKDSFTRTPDPRELEILKTVKTEITGFLIQAWPDNWTDLH 418

RESULT 6
TVCHLV

epidermal growth factor receptor precursor - chicken

N;Contains: protein-tyrosine kinase (EC 2.7.1.112) erbb

C;Species: Gallus gallus (chicken)

C;Date: 28-Feb-1986 #sequence_revision 05-May-1995 #text_change 04-Feb-2000

C;Accession: A27720; A00643

R;tax, I.; Johnson, A.; Hawk, R.; Sap, J.; Bellot, F.; Winkler, M.; Ullrich, A.; Vennstr

Mol. Cell. Biol. 8, 1970-1978, 1988

A;Title: Chicken epidermal growth factor (EGF) receptor: cDNA cloning, expression in mou

A;Reference number: A27720; MUID:88261272; PMID:3260329

Qy 424 VFONLOWIRGRIHLNGAYSLTLOGLISWLGRLSLRELGLALIHNTHLCFVHTVPMD 483
Db 419 AFENLEIIRGRTQKQGFSLAVGLNITSLGLSLKEISDGVIIISGNRLCYANTINWK 478
Qy 484 QLRPNHQALLHTANPEDECVEGEGIACHQLCARGHCWGPGTQCVCNCSQFRLGQBCVEE 543
Db 479 KLFGTPTNQTKIMNRAEKDCKAVNHVCNPLCSSEGCGWPEPRDCVSCQNSRGRECEK 538
Qy 544 CRVLOGLPREYVNAHCLPCHPECCQONGSVTCFGEADQCACAHYKDPFCVACRPSG 603
Db 539 CNLIEGEPREFVENSEICQHPCEPLQAMNITCTGRPDNCIQCAHYIDGPHCVKTCAG 598
Qy 604 VKPDLVMPITWKPDEEGACOPFINTCHSCVDLDDKGPAPQORASPLTSIVAAVVGILL 663
Db 599 IMGENNTL-VWKYADANNVCHLCHANCTGCAGPGLOGCEVWSPGPKISFIATGIUGLL 657
Qy 664 VVVLGVVFGI-LIKRRQOKIRKYTMRLRLOETELVEPLTPSGAMPNQAQMRILKETLRK 722
Db 658 FIVV-VALGIGLPMRRRHIVRKTLRLQLQERELVEPLTPSGEAPNQAHLRIKETEFK 716
Qy 723 VKVLGSGAGTGYKGIWIDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVSPYV 782
Db 717 IKVLGSGAGTGYKGLWIPEGEKVKIPVAIKELREATSPKANKEILDEAYVMASVDNPHV 776
Qy 783 SRLLGICLTSTVOLTPYGLLDHVRNRCGLSQDILLNMCQIAKGMSTLEVDRLV 842
Db 777 CRLLGICLTSTVOLITQILMPYGLLDYVREKDNIGSOVLLNWCVOJAKGMYLEDRVLV 836
Qy 843 HRDLAARNVLKSPNHNKJITDFGLARLLDIDETEHADGKVPKIKWMALESILRRRPTHQ 902
Db 837 HRDLAARNVLKTPQHVKITDFGLAKLLGAEKEEYHAEKKVPKIKWMALESILHRIYTHQ 896
Qy 903 SDVWSYGVTVWELMTGAKPYDGI PAREIPDLEKGERLPDPPICITIDVYMINVKCMID 962
Db 897 SDVWSYGVTVWELMTGSKPYDGI PASDISILEKGERLPDPPICITIDVYMINVKCMID 956
Qy 963 SECRPFRELVBSEFMSARDPQRFVVIQ-NEDLGASPLDSTFYRSLLEDDEDDGLVDAAE 1021
Db 957 ADSRPFRELILFESQVARDPQRYLVIQGERMHLSPDTSNFYRALMBEEDMEDVDAD 1016
Qy 1022 EYLVPOQGFPCPDPAFGAGGMVHRRSSSTRSGGDLTLGLPESEEAAPRSLAPSEGA 1081
Db 1017 EYLVPOQGF-----NSPST-----SRTLLSSLSA 1042
Qy 1082 GSDVFDGDLGMGAAGKLSLPHDPSPLQRYSEDPTVPLPSET--DGYVAPLTCSPQPEY 1139
Db 1043 TSN-----NSTVACINRNGSCRKVEDAFQRYSSDPTCAVTEDNIDDAFL-----VPPEY 1092
Qy 1140 VNQPDVVRPOPSPREGPLPAAPAGATLERAKTLPKNGVVKDVFAGGAVENPEYL-T 1198
Db 1093 VNO-SVPKRPAGSVQNPVYHNQDLHP-----AFGRDLHYQN--PHSNAVGNPEYLT 1141
Qy 1199 PQGGAAPPHPPPAFPAFDNLYYWDQ-----DP-----PERGAPSTPTFGTPT 1242
Db 1142 AQ-----PTCLSSGFSNPAWIQKSHQMSLNDPNYQQDFFPKETKPNKIFKG-PT 1191
Qy 1243 AENPEYLGIDVP 1254
Db 1192 AENAEYLRVAPP 1203

A:Accession: A27720
A:Molecule type: mRNA
A:Residues: 1-1223 <LAX>
R:Nilsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M.
Cell 41, 719-726, 1985
A:Title: C-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro
A:Reference number: A00643; MUID:85228222; PMID:2988784
A:Accession: A00643
A:Molecule type: mRNA
A:Residues: 585-1223 <NIL>
A:Cross-references: GB:M10066
C:Genetics:
A:Gene: erbB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: alternative splicing; ATP; autophosphorylation; glycoprotein; growth factor
specific protein kinase
F:1-30/Domain: signal sequence #status predicted <SIG>
F:31-1223/Product: epidermal growth factor receptor #status predicted <MAT>
F:31-654/Domain: extracellular #status predicted <EXT>
F:81-307/Domain: EGF receptor extracellular domain repeat <EE1>
F:397-610/Domain: EGF receptor extracellular domain repeat <EE2>
F:655-677/Domain: transmembrane #status predicted <TM>
F:678-1223/Domain: intracellular #status predicted <INT>
F:719-984/Domain: protein kinase homology <KIN>
F:727-735/Region: protein kinase ATP-binding motif
F:136,202,280,361,370,422,575,580,615,635/Binding site: carbohydrate (Thr) (covalent) #
F:192,650/Binding site: carbohydrate (Ser) (covalent) #status predicted
F:687/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F:754/Active site: Lys #status predicted
F:1100,1183,1208/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #stat
Query Match 45.9%; Score 3123.5; DB 1; Length 1223;
Best Local Similarity 48.7%; Pred. No. 2e-120;
Matches 632; Conservative 175; Mismatches 345; Indels 145; Gaps 25;
QY 8 RWGLLLALLPGAA-----STQVCTGTDMLRLPASPETHDMLRHLHYGCGVQVGNLE 61
DB 13 RGAVALVLLGLGVALCSAVEBEKVCQGTNNKLTQLGHVEDHFTSLQRMNVCVLSNLE 72
QY 62 LTYLPTNASLSFLQDIOEGVGYVLIHNOVRQVPLQRLRIVRGTLQPLDNDYVALAVLDNGD 121
DB 73 ITVEHNRDLTFLKTIQEVAGYVLIAMVDVIPLENLQIIRGNLVNDSFALAVLSNYH 132
QY 122 PLANNTPVTGASPGGLRELQRLSRLTEILKGVLIQRLNLPOLCYODTILWKDIFHKNQAL 181
DB 133 -MNTQT-----GLRELPMKRLSELINGVKVINSNNPKLNCMDTVLWNDIIDTSRK-PL 182
QY 182 TLID-TWRSRACHPCSPMGKSGRCWGESSEDCSLRTVTCAGGCA-RCKGPLPTDCHEQ 239
DB 183 TVLDFASNLSSCPKCHPNCTEDHCWAGAGQNCQTLTKVICAQCGRCGRGKVPSCDCHNQ 242
QY 240 CAAGCTGPKHSDCLACLFHNSHGICELHCPALVYNTDTFESMPNPEGRVTFGASCVTAC 299
DB 243 CAAGCTGPRSDCLACKFRDDATCKDTCPLVLYNPITYQMDVNPBGKYSFGATCVREC 302
QY 300 PYNLSTDVGSCTILVCPHLHQEVTAEDGTORCEKCKSPCARVCVGLGMEHLREVRVTS 359
DB 303 PHNVVVDHSGVRSNCNTDYEV-EENGVRKCKCDGLCKSVKNGVIGIGELKGLSINAT 361
QY 360 NIOEFACCKIPGLAPLPSFGDPDASNTAPIQEOQLQVETLEETITGYLTYSAMPDSL 419
DB 362 NIDSFNCKTNGDVSLPVAFLGDAFTKTLPLDPKLDVFTVKEISGLFLLIQAWPDNA 421
QY 420 PDLVSFONLQVIRGILHNGAYSILTQGLISGLRLSRLSELGSLALIHNTLHLCFVHT 479
DB 422 TDLYAFENLEIRGTQKHQGYSLAVNLKIQSLGRLSKLEISDGDIAIMKNKLCYADT 481
QY 480 VPWDQLFRNPQALLHTANPEPECVGEGGLACHQLCARGHCWGPQTQVCNCSQFLRGQE 539
DB 482 MNWRSUFATOSQTKIIQENKNDCTADRVCDPLCSDCVCGGFGPFCFSFRQKE 541
QY 540 CVEECRVQLGPREYNARHCLPCHPECQPNQ---SVTCFGEADQCVACAHYKPPFC 596

DB 542 CVKQCNIQGEPRFERDSKCLPCHSECLVQNSTAYNTTCSGPGPDHCKCAHFDIPGPHC 601
QY 597 VARCPGVKPDLSYMPIMWKPDPBEGACQPCPINCSTHSCVDLDDKGPAPAEORASPLTSIVS 656
DB 602 VRACPAVLGENDTL-WVKYADANAVCOLCHPNCTRGCKPGLEGCP---NGSKTFSIAA 657
QY 657 AVV-GILLVVVGVVGFILIKRQQKIRKYTWRRLLQETELVEPLTPSGAMPNQAQMRIL 715
DB 658 GVVGGLLCLVVGIGIGLYLRRR-HIVKRTLRLQLQERELVEPLTPSGEAPNQAHLRIL 716
QY 716 KETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAKVLRNTSPKANKELIDEAYVMA 775
DB 717 KETEFKVKVVLGSGAGFTVYKGLWIPDGENVKIPVAKVLRNTSPKANKELIDEAYVMA 776
QY 776 GVGSPVSVRLGLICTSTVQLVTLQMPYGCCLDHDVRENRGLSGDQLLNMCMQIAGMSY 835
DB 777 SVDNPHVCHLLGLICTSTVQLITQLMPYGCCLDHYREHKDNTGSGVLLNWCQIAGMNY 836
QY 836 LEDVRLVHRDLAARNVLVKSNNHVKITDFGLARLLDIDETEHADGGKVPKIKWMALESIL 895
DB 837 LEERRLVHRDLAARNVLVKTPOHVKITDFGLAKLGADEKEYHAEGKVPKIKWMALESIL 896
QY 896 RRRFTHQSDVWSYGVTVWELMTFGAKYDGIIPAREIPDLLEKGERLPOPPICTIDVYIM 955
DB 897 HRIYTHQSDVWSYGVTVWELMTFGSKPYDGIIPASEISSVLEKGERLPOPPICTIDVYIM 956
QY 956 VKCWMIDSCRRFRRLVSEFRSMARDPCRFVVIQ-NEDLGASPLDSTFYRLSLDDDDM 1014
DB 957 VKCWMIDASRRFRLIAFAFNMADPPRYLQODERHLPSPPTDSKYFTLMBEEDM 1016
QY 1015 GDLVDAEYLVFQQGFCFDPAPGAGGMVHRHRSSTRSGGDLTLGLEPSEEEAPRSP 1074
DB 1017 EDIVDAEYLVFQHGFF-----NSPST-----SRTP 1042
QY 1075 L-----APSEGAGSDVFDGDLGMAAKGLOSLPTHDPSPLOYSEDPVTPLPSET--DGY 1127
DB 1043 LLSLSATSNNSATNCID-----RNGQGHFVREDSEFVQSYSSDPTGNFLEESIDDFG 1094
QY 1128 VAPLTCSPQEVYNODVVRPQPPSPREGPLPAARPAGATLERAKTLSPCKNGVVKDVF-- 1185
DB 1095 L-----PAPEVYQ--LMPKPS-----TAMVQNLVYNNISLT 1125
QY 1186 -----AFGAVENPEYLTPOGGAAPQHPPPAFSPAFDNLVYWDQ----- 1225
DB 1126 AISKLPMDSRYSQNSHSTAVDNPEYL-----NTNQSPKLVTFESSFYQSGNHQIN 1177
QY 1226 -DPPE-----RGAPPSTFKGTPTAENPEYLGLDVP 1254
DB 1178 LDNPDYQDFLPLNETKPNGLKVPAAENPEYLRVAAP 1214
RESULT 7
A47253
Epidermal growth factor receptor, HER4 - human
C:Species: Homo sapiens (man)
C:Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 18-Jun-1999
C:Accession: A47253
R:Flouman, G.D.; Culoucou, J.M.; Whitney, G.S.; Green, J.M.; Carlton, G.W.; Foy, L.; Ne
Proc. Natl. Acad. Sci. U.S.A. 90, 1746-1750, 1993
A:Title: Ligand-specific activation of HER4/p180erbB4, a fourth member of the epidermal
A:Reference number: A47253; MUID:93189574; PMID:8383326
A:Accession: A47253
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1-1308 <PLO>
A:Cross-references: GB:L07866; NID:g337359; PID:AAB59446.1; PID:g337360
A>Note: sequence extracted from NCBI backbone (NCBIP:126842)
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; growth factor receptor
F:716-981/Domain: protein kinase homology <KIN>
F:724-732/Region: protein kinase ATP-binding motif

Db 123 YQK-NPSSP--DYQVGLKQLQLNLTEILSGGVKSHNPLLCNVETINWWDIVDKTSNP 179
Qy 180 ALTLLIDNRACRCHPCSMCKSGRCWGESSEDCOSLTRTVCAGGC-ARCKGPLPTDCCHE 238
Db 180 TMLNLI PHAFERQCKQCDHGCWNGSWAPGPGCHQKFTKLLCAEQCNRRCRGPKPIDCCNE 239
Qy 239 QCAAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCUTA 298
Db 240 HCAGGCTGPRATDCLACRDNDDGCTCKDTPPPKXIYDIVSHQVVDNPNIRKTYTGAACVKE 299
Qy 299 CPVNYLSTDVSGCTLVCLPLHNOEYTAEDGTORCEKSKPCARVCYGLGMEHLREVRVATS 358
Db 300 CPSNVVYTE-GACVRSACSAGMLEVD-ENGRSKCPDGVCPKVCDDGIGISLNTIAVNS 357
Qy 359 ANIQERAGCKKIFGSLAFLESFDDPASNTAPLQPELOQVFTLEBITGYLISAWPDS 418
Db 358 TNIRSFNCTKINGDIILNRNSPEGDPHYKIGTMDPEHLNLTIVKSEITGYLVIMWPFEN 417
Qy 419 LPDLSVFONQVIRGRILHNGAYS-LTLOGLGISWGLRSLRELGLSGLALIHNTHLCFV 477
Db 418 MTSLSVFONLEIIRGTRTTFGRGFSFVVQVRHLQWLGLRSLKEVSAGNVILKNTLQLRYA 477
Qy 478 HTVPWDLFRNPHQALLHTANRPEDECVGEGLAGCHQLCARGHCWGPCTOCVNCQSLRG 537
Db 478 NTINWRLLFRSEDSIYDART-----ENQTCNECEDGCV-POPTMCVSLCHVDRG 529
Qy 538 QECVECRVLQGLPREYVNAHCLPCHPEQOPQNGSVTCFGEADQCVACAHYKDDPPFCV 597
Db 530 GRCVASCNLLQGEPREAQVGRVCQHQECLVTDLSLTCYGPGRANCKSAHQDGGQCI 589
Qy 598 ARCPGVKPLSYMPYIKWPFDEBAGCQPCINCTHSCVLDLDDKGCAPABQASPLTSIVSA 657
Db 590 PRCPHGILGSDTL-IWKYADKMGQCPCHQNCCTGCGSLGCRGD-IVSHSSLAVAL 647
Qy 658 VVGILLVVLGVVFGILLKEROQKIRKYTWRRLLLOETELVEPLTPSGAMPNQAMRLKE 717
Db 648 VSGLLIIVALLIVLVLVRLRRIR-KRRTIRCLLQEKELVEPLTPSQAPNQAFRLIKE 706
Qy 718 TELRKVKLGSGAGTGYKGIWIIPDGENVKIPVAIKVLRNTPSPKANKEILDEAYVMAGV 777
Db 707 TEFKDKRVLGSGAGTGYKGLMNPDDGENIRIPVAIKVLRNTPSPKANKEILDEAYVMAGV 766
Qy 778 GSPVSRLLGICLTSTVQLVTQMLPYCCLLDHVRNGRGLGSQLNWCQIAKMSYLE 837
Db 767 DHPHVCRLGICLTSAVQLVTQMLPYCCLLDYVRQHOERICGOWLLNWCQIAKMSYLE 826
Qy 838 DVLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESILR 897
Db 827 ERHLVHRDLAARNVLLKNPNHVKITDFGLSKLLTADKEYQADGGKVPKIMMALESILQW 886
Qy 898 RFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYIMVK 957
Db 887 TYTHQSDVMSYGVTVWELMTFGSKPYDGIIPAKIASVLENGERLPQPPICITIEVYIMLK 946
Qy 958 CWMIDSECRPRFRELVESEFMRMARDPQFVVIQNEIDLGASPLDSTFYRSLLEDGDDGDL 1017
Db 947 CWMIDPSRRPRFRELVEGFQMARDFRYLVIQG---NLPSLSDRRLFSRLSSDD--DV 1001
Qy 1018 VDAEYLVPQGGFCPPDAPGAGGVVHRRSSSTRSGGGLDTLIGLEPSEEAAPRSLAP 1077
Db 1002 VDAEYLLPYKRI-----NRQGS-----EPCIP 1024
Qy 1078 SEGAGSDVFDGDLGMAAGKGLQLPHTDPSFLORYSBDPTV-PLPSSTDGVAPLTCSPQ 1136
Db 1025 PTGH-----PVRENSITLRNSTDPTQNALEKDLGH----- 1055
Qy 1137 PEYVQDVRPQP-----PSPRE-----GPLP-AARPAGATLERAKTLSPQKNGVVKD 1183
Db 1056 -EYVNPQGETSRLSDIYNPNVEDLTDGWPVSLSSQEAETFSRPEYLNTQNSL--- 1111
Qy 1184 VPAFGGAVENPEYLTPOGGAAPQHPHPPAPSPAFDNLVYWDQDPPEGAPSPFTKPTTA 1243
Db 1112 PLVSSGMDPDY---QAG-----YQAAF-----LPQTGALTGNGMFLPAA 1149

Qy 1244 ENPEYLG 1250
Db 1150 ENLEYLG 1156

RESULT 9

A36223

Kinase-related transforming protein (erbb3) (EC 2.7.1.-) precursor - human
C;Species: Homo sapiens (man)

C;Date: 04-Oct-1991 #sequence_revision 13-Jan-1993 #text_change 17-Nov-2000

C;Accession: A36223; I59164

R;Kraus, M.H.; Issing, W.; Miki, T.; Popescu, N.C.; Aaronson, S.A.

Proc. Natl. Acad. Sci. U.S.A. 86, 9193-9197, 1989

A;Title: Isolation and characterization of ERBB3, a third member of the ERBB/epidermal g

A;Reference number: A36223; MUID:90083234; PMID:2687875

A;Accession: A36223

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-1342 <KRA>

A;Cross-references: GB:M29366

R;Plozman, G.D.; Whitney, G.S.; Neubauer, M.G.; Green, J.M.; McDonald, V.L.; Todaro, G.J

Proc. Natl. Acad. Sci. U.S.A. 87, 4905-4909, 1990

A;Title: Molecular cloning and expression of another epidermal growth factor receptor-re

A;Reference number: I59164; MUID:90311312; PMID:2164210

A;Accession: I59164

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: mRNA

A;Residues: 1-559, 'G', 561-957, 'F', 959-1063, 'G', 1065-1342 <RES>

A;Cross-references: GB:M34309; NID:G183990; PIDN:AAA35979.1; PID:G306841

C;Genetics:

A;Gene: GDB:ERBB3; HER3

A;Cross-references: GDB:119880; OMIM:190151

A;Map position: 12q13-12q13

C;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolog

C;Keywords: ATP; phosphotransferase

F;707-972/Domain: protein kinase homology <KIN>

F;715-723/Region: protein kinase ATP-binding motif

Query Match 35.7%; Score 2431.5; DB 2; Length 1342;

Best Local Similarity 40.7%; Pred.No. 3.7e-92;

Matches 533; Conservative 191; Mismatches 458; Indels 129; Gaps 32;

Qy 10 GLIALALPPGAA--STQVCTGTDMKRLPASPETHLMLRLHYQGQVQVGNLELYLPT 67

Db 11 GLIFSLARGSEVGNQAVCPGTLNGLSVGTDAENQYQTLVLYRCEVVMGNLEIVLTGH 70

Qy 68 NASLSFLQDIQEVQGVYLAHQVQVPLQRLRIVRGTLFEDNYALAVLDNGDPLNNTT 127

Db 71 NADLSFLQIREVTGYLVVAMNEFSTLPNLRVVRGTQVYDGKFAIFVM-----LYNNT 125

Qy 128 PVTGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYODTILMKDIPHKKNQALTLIDTN 187

Db 126 ----NSSHALRQLRLTQLTEILSGGVYIEKNOKLCHMDITDWRDIVDRD---AEIVVKD 178

Qy 188 RSRACHPCSPMCKSGRCWGESSEDCOSLTRTVCAGGC-ARCKGPLPTDCHEOCAAAGCTG 246

Db 179 NGRSCPPCHEVCKG-RCWPGSSEDCQTLTKTICAPQCNHGCFGNPQCHDECAGCGSG 237

Qy 247 PKHSDCLACHFNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACPYNYLST 306

Db 238 PQDTCDFACRHFNDSGACVPCPQPLVYNKLTFOLEPNPHTTKYQGVGVASCAPHNFV-V 296

Qy 307 DVGSCTLVCLPLHNOEYTAEDGTORCEKSKPCARVCYGLGMEHLREVRVATSNIQEPAG 366

Db 297 DOTSCVCRACPPDKMEVD-KNGLKMCPECGGLCFKACEGTGSG--SRFQTVDSNIDGFVN 353

Qy 367 CKKIFGSLAFLESFDDPASNTAPLQPELOQVFTLEBITGYLISAWPDSLPDLSVFO 426

Db 354 CTYKILGNLDFLTGLNGDPWHKIPALDPEKLVNFRVTRITGYLNIQSPPHMHPNFSVES 413

Qy 427 NLQVIRGRILHNGAYS-LTLOGLGISWGLRSLRELGLSGLALIHNTHLCFVHTVPWQOL 485

Db 414 NLTTIGRSYLRNFGSLIMKLNVTSLGFRSLKEISAGRIYISANRQLCVHSLNWKV 473
Qy 486 FRNPHQALLHTA-NRPEDECVEGLACHOLCARGHGWPGTQCVCNCSQFLRGQECVEEC 544
Db 474 LRGTTERELDKHNRPRDCVAEKVCDPLCSCGGCGPGGQCLSCNYSRGVCVTHC 533
Qy 545 RVLGGLPREYVNRHCLPCHPECPQNGSVTCFPEADQCACAHYKDPFPCVARCPGV 604
Db 534 NFLNGEPREFAHEAECSCHPECPMEGTATCNGSGSDTCAOQCAHPRDGHVCVSSCPGV 593
Qy 605 KPDLSYMPIKFPDEBACQPCPINCSTHSCVDLDDKCPABORA-----SPLTSIVSAVG 660
Db 594 LG--AKGPIKYPDVQNECRCHENCCTGCKGPELODCLGQTLVLKGLTHLWALVIAG 651
Qy 661 ILLVVLGVVFGILIKRQOKIR-KYTMRLLOSTELVEPLTPSGAMPNOAQMRILKETE 719
Db 652 --LVVIFMVLGGTFLYWRGRIQKRAMRYRLGESIEPLDPS-EKANKVLARIFKETE 708
Qy 720 LRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKELDEAYVMAGVS 779
Db 709 LRKLVKLVGSGVGVYKGVWIPGESIKIPVCIKVIEDKSGRQSFQAVTDHMLAIGSLDH 768
Qy 780 PYVSRLLIGICLTSTVQLVTQMLPVGCLLDHVRNRLGSDLLNWCMTAKGMSYLEDV 839
Db 769 AHVRLGLCLPGSSLOLVTOYPLGLSLDHRQHRGALGPQLLLNMGVQIAKGMYLEEH 828
Qy 840 RLVRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMMALRESILRRF 899
Db 829 GWHRNLAARNVLKSPQVQADGVADLLPPDDKQLLYSEAKTPIKMMALRESIHFGY 888
Qy 900 THQSDVMSYGVYVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVVMVWKW 959
Db 889 THQSDVMSYGVYVWELMTFGAEPYAGRLAEVDPDLLEKGERLAQPOICTIDVVMVWKW 948
Qy 960 MIDSECPRELVSERWARDQRFVWIONEDLGA-----SPLDSTFYRSLLEDDMDG 1016
Db 949 MIDENIRPTEKELANEFTRWARDPRYLVIKRES-GPGIAPGEPHGLTKLEEEVELEP 1007
Qy 1017 LVDAEYLVPOQGFCDPAPGAGGMVHRHRSSTRSGGLTLGLEP-SEEEAPRSL 1075
Db 1008 ELDDLDELEED-----NLATTLGSLSLPVGTLNRRFGSOLL 1048
Qy 1076 APSEGAGSDVFDGLGMAAGLQSLPTH-PSPLQRYSEDPTVPLP-----SETDGYV 1128
Db 1049 SPSSGY-MPMNOGNLGSQESAVSGSSRCPVSLH-----PMRPGCLASSESGHV 1101
Qy 1129 A-----PLTCSPOPE-----YVNPQDVRPOPSPREP-----L 1157
Db 1102 TGSEAELOEKVSMCRSRSRSPRGRDSAYHSQRHSLLTPTVPLSPGLEEEDVNGYV 1161
Qy 1158 PAARPAGATLERAKTLP-GKNGV-----KOVFAFGGAVENPEYLTPOGGAPOPHPP 1210
Db 1162 PDTHLKTSPSREGTSSVGLUSSVLGTEEDED-----EVEYNNRRRRHSP-PHPP 1212
Qy 1211 PAFSPAFDNLVYWD-----QDPPERGAPPSTFKTPTAENPEYL 1249
Db 1213 RPSLEELGVEYMDVGSLSASLGTSQCLHPVPIMPTAGTTPDEYEM 1263

RESULT 10

JC4387
epidermal growth factor receptor homolog precursor - rat
N:Alternate names: ErbB3 protein; HER3 protein
C:Species: Rattus norvegicus (Norway rat)
C>Date: 17-Jan-1996 #sequence_revision 19-Apr-1996 #text_change 13-Nov-1998
C:Accession: JC4387
R:Hellmeyer, N.J.; Kim, H.H.; Greaves, C.H.; Sierke, S.L.; Koland, J.G.
Gene 165, 279-284, 1995
A:Title: Cloning of the rat ErbB3 cDNA and characterization of the recombinant protein.
A:Reference number: JC4387; MUID:96096535; PMID:8522190
A:Accession: JC4387
A:Molecule type: mRNA
A:Residues: 1-1339 <HEL>

A:Cross-references: GB:U29339; NID:G915389; PID:G915390
A:Experimental source: liver
C:Note: The authors translated the codon AAC for residue 369 as Thr and GGT for residue 1
C:Comment: This protein is a functional heregulin receptor that transduces signals to the
C:Genetics:
A:Gene: ErbB3
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolog;
C:Keywords: ATP; growth factor receptor; liver; phosphoprotein; transmembrane protein
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-1339/Product: epidermal growth factor homolog #status predicted <MAT>
F:640-659/Domain: transmembrane #status predicted <TM>
F:705-970/Domain: protein kinase homolog <KIN>
F:713-721/Region: protein kinase ATP-binding motif
F:939,1051,1156,1194,1196,1219,1257,1259,1273,1286,1325/Binding site: phosphate (Tyr) (cc
Query Match 34.4%; Score 2346.5; DB 2; Length 1339;
Best Local Similarity 40.8%; Pred. No. 1.1e-88;
Matches 523; Conservative 171; Mismatches 434; Indels 155; Gaps 34;
Qy 3 LAALCRWGLLLALLPPGAA---STVCTGTDMKRLPASPETHDMLRLHYLQCGQVVOGN 59
Db 7 LQVLC-----FLSLARGSEMGNSQAVCPGTNLGLSVTGADNQYQTLYKLYKECEVMGN 62
Qy 60 LEITYLPTNASLFLQDIOEVQGYVLIHNRQVPLQRLRIVRGTQLFEDNVALVLDN 119
Db 63 LEIVLTGHNADLSFLQWIREVTAYVLVAMNEFSVLPLNLRVVRGTQVYDGFKAIFVM-- 120
Qy 120 GDPNLTPTVTGASPGGLRELQRLSLEITLKGVLQRLNPOLCYQDTILWKDIFHKNNQL 179
Db 121 ---LNTNT---NSSHALRQLKFTQLTEILSGSVYIEKNDKLMCHMDTIWRDILVRVR--- 170
Qy 180 ALTLIDTRNSRACHPCSPMKSGSRGSESDCSLTRVTWCAGGC-ARCKGPLPTDCCHE 238
Db 171 GAEIVVKNANGANCPCHEVCKG-RCWGPGPDCCQLITKLTICAPQNGRCFEPNQCCHD 229
Qy 239 QCAAGCTGPKHSDCLACLHFNHSGICELHCPALVYNTDTTFESMPNPEGRYTFGASCVTA 298
Db 230 ECAGGCGQPDTCFACRRFNDSGACVPCPEPLVYNKLTFLQEPNPHTKYQYGGVCVAS 289
Qy 299 CPNYLSTDVGSCTLYCPLHNOEVTAEADGTORCEKSKPCARVCYGLGMEHLREVRVTS 358
Db 290 CPNHFV-VDTTCFVRACPPDKMEVD-KHGLKMPCECGGLCPKACEGTSGS--SRYQTVD 345
Qy 359 ANIQEPAGCKKFGSLAFLEPESFDGDPASNTAPLOEQVPELLEIETGLYVISAAPDS 418
Db 346 SNIDGFVNCTKILGNLDLITGLNVDPWKIPALDEKLNVRFTVREITGLYNIQSWPPH 405
Qy 419 LPDLSVFONLQVIRGRILHNGAYS-LTQGLGISWLGRLSLRGLSGLALIHHTHLCFV 477
Db 406 MHNFSVFNLTITGGRSLYNRGFSLLIMKNLVNLSLGRSLKEISAGRVVISANQQLCVH 465
Qy 478 HTVPMDQLFRNPHQALLHTA-NRPEDECVEGLACHOLCARGHGWPGTQCVCNCSQFLR 536
Db 466 HSLNLTLLRGLPSEERLDIKYDRPLGCELAEGKVCDCPLCSCGGCGPAPGQCLSCNYSR 525
Qy 537 GQECVEECRVLOGLPREYVNRHCLPCHPECPQNGSVTCFPEADQCACAHYKDPFPC 596
Db 526 EGVVTHCNFLOGEPREFVHEAQCFCHPECLPMEGTSTYNGSGSDACARCAHFRGPHC 585
Qy 597 VARCPGSKVPDLSYMPIKFPDEBACQPCPINCSTHSC--VDLDDKGCFAEQRASPLTSI 654
Db 586 VNSCPHGILG--AKGPIKYVPOAQNCRFCHENCCTGCGNPELQDCLGQAEVLMSKPHLV 643
Qy 655 VSAVGVILLVVLGVVFGILIKRQOKIR-KYTMRLLOSTELVEPLTPSGAMPNOAQMR 713
Db 644 IAVTVG--LAVILMILGSGFLYWRGRIQKRAMRYRLGESIEPLDPS-EKANKVLAR 700
Qy 714 ILKETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKELDEAYV 773
Db 701 IFKETELRKVLGSGVFGTVHGIWIPGESIKIPVCIKVIEDKSGRQSFQAVTDHMLA 760
Qy 774 MAGVGSPPYVSRLLIGICLTSTVQLVTQMLPVGCLLDHVRNRLGSDLLNWCMTAKGM 833

Db 761 VGSIDRAHIVRLGLCPGSSQLVQYPLGLSLLDHVKQHQHRELTPQLLLNWGVQIAKGM 820
QY 834 SYLEDVRLVHRDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGGKVPPIKWMALLES 893
Db 821 YYLEEHSVMVRDLAARNVLMKSPSQVQVADFGVADLLPPDDKQLLHSEAKTPIKWMALLES 880
QY 894 ILRRRTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYM 953
Db 881 IHFGKYTHQSDVWSYGVTVWELMTFGAEPYAGRLAEIPDLLEKGERLAQPOICTIDVYM 940
QY 954 IMVKCWMIDSECRPRELSEFSRMARDPQRFVWJQNEQDLGPASPLDSTFYRSILLEDD 1013
Db 941 VMVKCWMIDENIPTKELANETRWARDPRVILVIKRAAS-GRCTP--PAEFSVITTK 997
QY 1014 MGDVDAEYLVPOQGFPCPDPAAGAGGMVHRRHSSTRSGGDLTLGLEPSEE----- 1068
Db 998 L-----QEALEPEL-----DLDLDLEAESEGLATS 1023
QY 1069 -----EAPRSLAPSEG-----AGSDVDFDGLGMGAAGKLGSLPETHD 1105
Db 1024 LGSALSPLPTGTLTRPGSQSLSPSSGYMPPMNOSSILGEACLDASVILGGRQFGRPISLH- 1082
QY 1106 PSLQRYSEDPVPLPSETDGYV-----APL-----TC-----SPOPE-----VYNQPDV 1145
Db 1083 PIPRGR-----PASESEGHWTGSEAELOEKVSVCRSRSRSPRPRGDSAYHQSQRHS 1135
QY 1146 RPQPPSPREGP-----LPAARPAAGATLERAKTLSP-GKNGVV-----KDVFAF 1187
Db 1136 LLTPVTPLSPPGLEEDGNGVMPDTHLRCASSRRECTLSVGLSSVLGTREDEDED----- 1191
QY 1188 GGAVENPEYLTPOGGAAPOPHPP 1210
Db 1192 -----EEYENMRKRGSP-PRPP 1209

RESULT 11
TVFVLV
protein-tyrosine kinase (EC 2.7.1.112) erbB - avian leukosis virus
N;Contains: amino end of gag protein; env protein fragment; protein-tyrosine kinase
C;Species: avian leukosis virus, ALV
C;Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 11-Jun-1999
C;Accession: B00643; A00643
R;Nielsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M.
Cell 41, 719-726, 1985
A;Title: C-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro
A;Reference number: A00643; MUID:85228222; PMID:2988784
A;Accession: B00643
A;Molecule type: mRNA
A;Residues: 1-698 <NIL>
A;Cross-references: GB:M10066; GB:M13881; NID:G211749; PIDN:AAA48763.1; PID:G211750
A;Note: in Genbank entry CHKERBBF, release 109.0, the source is designated as Gallus gal
C;Comment: This protein is synthesized as a gag-env-erbB protein.
C;Genetics:
A;Gene: gag-env-erbB
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific P
F;1-6/Product: gag protein (fragment) #status predicted <GAG>
F;7-59/Product: env protein (fragment) #status predicted <ENV>
F;60-698/Product: protein-tyrosine kinase erbB #status predicted <ERB>
F;194-459/Domain: protein kinase homology <KIN>
F;202-210/Region: protein kinase ATP-binding motif
F;229/Active site: Lys #status predicted

Query Match 25.98; Score 1766.5; DB 1; Length 698;
Best Local Similarity 52.28; Pred. No. 2.6e-65;
Matches 374; Conservative 80; Mismatches 137; Indels 125; Gaps 18;

QY 578 GPEADQCVACHYKDPFCVARGCPGVKPDLSYMPYKWKFPDREGACPCPINCTHSCVDL 637
Db 60 GP--DHCMKCAHFDGPHCVKACPAVGLGENDL-VWKYADANAVCOLCHPNCTRCKGP 116
QY 638 DRKGCBAEQRASPLTSTVSNAV-GILLVVGLVGVFGILIKRQOKIRKYTMRLLOTEL 696

Db 117 GLEGCP---NGSKTPSIAAGVVGGLLCLVVVGLGIGLYLRRR-HIVRKETLRLLOEREL 172
QY 697 VEPLTPSGAMPNQAOQMRILKETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLR 756
Db 173 VEPLTPSGEAPNQAHRLILKETEFKVKVKVLGSGAGFTVYKGLWIPBGEKVKIPVAIKELR 232
QY 757 ENTPSKANKEIIDEAYVMAGVSPVYSRLIGTCLSTVOLTQVLTQMPYGCGLLDHVRNRR 816
Db 233 EATSPKANKEIIDEAYVMASVDNPHVCRLLGTCLTSTVQLITQMPYGCGLLDYIREHKON 292
QY 817 LGSQDLNWCWMOIAKMSYLEVDVRLVHRDLAARNVLKSPNHNKITDFGLARLLDDIDE 876
Db 293 IGSQVLLNWCVOIAKMNVLLEERLVHRDLAARNVLKTPQHVKITDFGLAKLLGADEKE 352
QY 877 YHADGSKVPIKWMALLESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLE 936
Db 353 YHAEGSKVPIKWMALLESILHRYTHQSDVMSYGVTVWELMTFGSKPYDGIIPASEISSVLE 412
QY 937 KGERLPPOPICTIDVYIMVKCWMIDSECRPRELSEFSRMARDPQRFVVIQ-NEDLG 995
Db 413 KGERLPPOPICTIDVYIMVKCWMIDADSRPRKRELIAEFSKWARDPPRYLVIOGDERMH 472
QY 996 PASPLDSTFYRSILLEDDMDGLVDAEYLVPOQGFPCPDPAAGAGGMVHRRHSSTRSG 1055
Db 473 LPSPTDSKPYRTLMSEEDMEDIVDADEYLVPHQGF-----NSPST--- 513
QY 1056 GGDLTILGLEPSEEEAPRSPL-----APSEGAGSDVDFDGLGMGAAGKLGSLPETHDPSPLQ 1110
Db 514 -----SRTPLSSLSATSNSATNCID-----RNGQHPVREDSFVQ 550
QY 1111 RYSEDPVPLPSET--DGTVAPLTCSPQEPYVNPQDVRPQPPSPREGPLPAARPAGATLE 1168
Db 551 RYSSDPTGNFLEESIDDGFL-----PAPEYVNO--LMPKXPS----- 585
QY 1169 RAKTILSPGKNGVYKQVF-----AFGGAVENPEYLTPOGGAAPOPHPPAF 1213
Db 586 -----TAMVQNIYNISLTAISKLPMDSRVQNSHSTAVDNPEYL-----NTNOSPLA 633
QY 1214 SPADFNLVYWDQ-----DPPE-----RGAPPSTFGTPTAENPEYLGLDVP 1254
Db 634 KTVFESSPYIQSGNHQINLDNPDYQDQFLPNETKENGILLKVPAAENPEYLRVAAP 689

RESULT 12
TVFVUH
protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus (strain H)
C;Species: avian erythroblastosis virus
C;Date: 18-Apr-1984 #sequence_revision 18-Apr-1984 #text_change 11-Jun-1999
C;Accession: A00644; A38022
R;Yamamoto, T.; Nishida, T.; Miyajima, N.; Kawai, S.; Ooi, T.; Toyoshima, K.
Cell 35, 71-78, 1983
A;Title: The erbB gene of avian erythroblastosis virus is a member of the src gene family
A;Reference number: A00644; MUID:84026539; PMID:6313229
A;Accession: A00644
A;Molecule type: DNA
A;Residues: 1-604 <YAM>
A;Cross-references: GB:K01216; NID:G209676; PIDN:AAA42400.1; PID:G209678
R;Debuire, B.; Henry, C.; Benaissa, M.; Biserte, G.; Claverie, J.M.; Saule, S.; Martin, J.
Science 224, 1456-1459, 1984
A;Title: Sequencing the erbA gene of avian erythroblastosis virus reveals a new type of
A;Reference number: A38022; MUID:84223957; PMID:6328658
A;Accession: A38022
A;Molecule type: DNA
A;Residues: 1-28, 'W', 30-139, 'P', 141-145, 'V', 147-152 <DEB>
A;Cross-references: GB:K02006
C;Genetics:
A;Gene: erbB
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific P
F;130-395/Domain: protein kinase homology <KIN>
F;138-146/Region: protein kinase ATP-binding motif
F;165/Active site: Lys #status predicted

Query Match 25.0%; Score 1703; DB 1; Length 604;
Best Local Similarity 52.2%; Pred. No. 8.9e-63;
Matches 360; Conservative 76; Mismatches 128; Indels 126; Gaps 16;
QY 587 CAHYKDPFPCVACPSGVKPDLSVMPYWKSPDEGACOPCPINCTHSCVLDKDGKPAEQ 646
DB 3 CAHFIDGPHCVKACPAVGLGENDTL-VRKYADANAVCOLCHPNCTRGCKPGLEGCP-- 58
QY 647 RASPLTSIVSAV-VILLVVLGVVFGILKRRQOKIRKVTMRLLQETELVPLTPSGA 705
DB 59 NGSKTPTSIAGVGGCLLVVGLGIGLYLRR-HIVRKTLRLRLQERLVEPLTPSGE 117
QY 706 MPNOAQRILKETELRKVKVYKVGSGAGFTVYKGIWIPDGENVKIPVAIKVIRENTSPKANK 765
DB 118 APNOAHLRIKETEFKVKVYKVGSGAGFTVYKGIWIPDGENVKIPVAIKVIRENTSPKANK 177
QY 766 EILDEAVVMAVGSPPYRILGICLTSTVOLVQLMPYGLLDHVRNRLGSDLLNW 825
DB 178 EILDEAVVMAVDNPHVCRLLGICLTSTVOLVQLMPYGLLDHVRNRLGSDLLNW 237
QY 826 CMQIAKMSYLEDLVRLVHRLDARLVKSPNVKIDTDFGLARLLDIDETEHADGKVP 885
DB 238 CVQIAKGMVLEERLVHRLDARLVKTPQHVKITDFGLAKLLGADKEHYHAEKGKVP 297
QY 886 IKWALESIILRRFTHQSDVMSYGVTVWELMTGAKPYDGIPIAREIPDLLEKGERLPQPP 945
DB 298 IKWALESIILRRFTHQSDVMSYGVTVWELMTGSKPYDGIPIAREIPDLLEKGERLPQPP 357
QY 946 ICTIDVYIMVKWIMIDSECRPRELVSEFSRMARDPQFVIO-NEDLGASPLDSTF 1004
DB 358 ICTIDVYIMVKWIMIDSECRPRELVSEFSRMARDPQFVIO-NEDLGASPLDSTF 1417
QY 1005 YRSILLEDDMDGLVDAEYLVPOQFPCDPAPGAGGMVHRRSSSTRSGGDLTLGLE 1064
DB 418 YRTIMEEDMEDVDAEYLVPOQGF-----NSPST----- 449
QY 1065 PSBEEAPRSP-----APSEGAGSDVEDGLGMGAAGLQSLTHPSPLOQRYSEDPVP 1119
DB 450 -----SRTPLLSLSATSNNSATNCID-----RNGQGHVPREDSFVORYSSDPTGN 495
QY 1120 LPSET--DGIVAPLTCSPQBYVNPQVRPQPSREGPLPAARPAGATLERAKTILSPGK 1177
DB 496 FLEESIDDDGL-----PAPEYVNO--LMPKPTSTAM----- 524
QY 1178 NGVYKDVFAF-----GGAVENPEYLTPOGGAAPQHPPPAFSPAF 1218
DB 525 --VQNIYNEFISLTAISKLPMDSRVQNSHSTAVDNPEYL-----NTNQSLAKTVFE 574
QY 1219 NLYYWDQDPPERGAPPSTFGTPTAENPEY 1248
DB 575 SSPYWIQSGNHQ-----INLDNPDY 594
RESULT 13
GQFE
epidermal growth factor receptor - fruit fly (Drosophila melanogaster)
N;Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB
C;Species: Drosophila melanogaster
C;Date: 17-Mar-1987 #sequence_revision 17-Mar-1987 #text_change 11-Jun-1999
C;Accession: A00640; A38021
R;Livneh, E.; Glazer, L.; Segal, D.; Schlesinger, J.; Shilo, B. Z.
Cell 40, 599-607, 1985
A;Title: The Drosophila EGF receptor gene homolog: conservation of both hormone binding
A;Reference number: A00640; MUID:85124611; PMID:2982499
A;Accession: A00640
A;Molecule type: DNA
A;Residues: 1-1330 <Liv>
R;Wadsworth, S.C.; Vincent III, W.S.; Billoreau-Wentworth, D.
Nature 314, 178-180, 1985
A;Title: A Drosophila genomic sequence with homology to human epidermal growth factor re
A;Reference number: A38021; MUID:85137938; PMID:2983232
A;Accession: A38021

A;Molecule type: DNA
A;Residues: 'A', 832-866, 'V', 868-943, 'QTPSLVK', <WAD>
A;Cross-references: EMBL:X02291; NID:G7922; PIDN:CAA26157.1; PID:G929565
C;Comment: This sequence is tentative because the introns have not been identified.
C;Genetics:
A;Gene: FlyBase:Egfr
A;Cross-references: FlyBase:FBgn0003731
A;Map position: 2 57p
A;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
F;1-732/Domain: extracellular #status predicted <EXT>
F;733-764/Domain: transmembrane #status predicted <TM>
F;765-1330/Domain: intracellular #status predicted <INT>
F;808-1072/Domain: protein kinase homology <KIN>
F;816-824/Region: protein kinase ATP-binding motif
F;122-300,324,363,518,688,695,700/Binding site: carbohydrate (asn) (covalent) #status pr
F;774/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F;843/Active site: Lys #status predicted
F;1181/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #status predicte
Query Match 24.3%; Score 1652.5; DB 1; Length 1330;
Best Local Similarity 29.9%; Pred. No. 2.2e-60;
Matches 413; Conservative 180; Mismatches 415; Indels 371; Gaps 39;
QY 80 VQGYVLIHNVQRVQLRQIRIVRGTFQLF-----EDNYALAVLDNGDPLNNTTPTVTGASP 134
DB 38 ITNYIVIGLDLIPCTLSYRLQIRGTLFSLSVVEEKYALFV-----TY 81
QY 135 GGLREQLRLSLEILKGGVLIQRNPOLCYQDTILWKIDIFHKNQLALTIDNRSRACHP 194
DB 82 SKMYTLEIPDLRDLVNGVGFHNNVNLCHMTIOWSEIVSNGTDAYNYDFETAPERCEPK 141
QY 195 CSPMGKSGCWSESSDDCSLRTVTCAGCA--RCKGPLPTDCCHCQCAAGCTGPKHSDC 252
DB 142 CHESCTHG--CWEGPKNCQKFKSLTCSQACGRCYGPKECHCFAGGCTGPTQKDC 200
QY 253 LACLHNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACPNYLSIDVGSCT 312
DB 201 IACKNFFDEAVSKEECPPMRKYNPTTYVLETPGKYAYGATCVKECP--GHLRLDNGACV 259
QY 313 LVCPLHNQEVTAADGTQRCCKSPCARVCYGLGMBHLREVRATVSANTOEPAAGCKKIFG 372
DB 260 RSCPDQKMDKGE-----CVPCNGPCPKTCPGVTVLH-----AGNIDSFRNCTVIDG 306
QY 373 SLAFLPESPDG--DPASNTA-----PLOPEQLQVETLEEITGYLYISAWPDSLPLSV 424
DB 307 NIRLDQTFSGQDVVYNTMGPRYIPLDPERREVSTVKETITGYLNIETHPQFNLSY 366
QY 425 FQNLQVIRGRILHNGAY-SLTLQGLGISWGLRSLRSLRELGLALIHNTHLCFVHTVPWD 483
DB 367 FRNLETHGRQLMESFAALAIVKSSLSYLEMRNLKQISSGVVIOHNRDLVCVSNIRWP 426
QY 484 QLFPRNPQALLHTANRPEDE----- 504
DB 427 AIQKEPEQKVWVNNENLRADLCGKFLTLISVQHNIIIMHIFAICREKNHLLGASVQRGLL 486
QY 505 ----- 504
DB 487 GSWHGSVPYLOELQFQWHLHRLRLWLYQVINSITQDKSNEHQLTDACYSFSPVPTSI 546
QY 505 -----VGEGLA-- 510
DB 547 ARYATQSLAMELEQITARSASMRHSKTLPAEGRQVPRWFLGVCA SARAGIAEPLAGR 606
QY 511 -----CHQICARGHCWGPPTQCVNCSQFLRQECVEECRVLOGLPREVY---NARHCLP 562
DB 607 AVCRKCHPLCELCTNYGHEQVCSKTKYKREQCETEC-----PADHYTDEQRECFQ 660
QY 563 CHPEQOPQNGSVTCGPEADOCVACAHYK-----DPPF-----CVARCPSGVK-PDL 608
DB 661 RHPEC---NG---CTGPGADDCSKRNFKLFDANETGYPYVNSTMTNCTSKCPLMRHVN 714
QY 609 SYMPIWKFPDEEGACQPCPCINCTHSCVLDLDDKGCAPAEQASPLTSIVSAVVGLLVVVLG 668

Db 715 QYTAIGPY-----CAASPPRSSKITANLD-----VNMIPIITGAVLVPTIC 755
QY 669 VFGI-LIKRQOKIRKYT--MRLLQETELVPLTPSGAMPNQAOIRILKETELARKVY 725
Db 756 ILCVVITYCQKQKAKKETVMTWALSGRESEPLRPSNIGANLCKLIVKDAELRGV 815
QY 726 LSGGAGTVYKGIWIDGENVKIPVAIKVURENTSPKANKEIIDEAYVMAGVGSPPVSR 785
Db 816 LMGAFGRVYKGVWPEGENVKIPVAIKELLKSTGAESSEFEFLREAYIMASBEHVLLKL 875
QY 786 LGICLSTVOLVQMPYGLLDHVRNRCGLSODLLANCMQIAKMSVLEDRVLVHRD 845
Db 876 LAVCMSQMLITQLPLGCLLDYVRNNRDKIGSKALLNWNSTQIAKMSVLEEKRLVHRD 935
QY 846 LAARNVLVK---SPNVHVKITDFGLARLLDIDETEHADGGKVPKWMALSIILRRFTHQ 902
Db 936 LAARNVLRLAGEDH---DFGLAKLLSSDSNEYKAAAGKMPKWLALCIRNVFTSK 991
QY 903 SDVWSGVTVWELMTGAKDYDGPAREIPDLLEKGERLPQPPICITIDVYIMVIMVKMID 962
Db 992 SDVWAFGVITWELTQGRPHENIPAKDIPDLIEVGLKLEQPEICSLDIYCTLLSCHWLD 1051
QY 963 SECRPFRELVSEFSRMARDPQRFVVIQEDLG--PASPLDSTFYRSLLEDD---DWGDL 1017
Db 1052 AAMRPTFKQLTTFABFARDPGRYLAIGDKFTRLEA-----YTSQDEKOLIRKLAPT 1104
QY 1018 VDAEYELVPOQGFPCPDPAAGAGVHHRSSSTRSGGDLTLGLEPSEERAP----- 1071
Db 1105 TDGSEAIKPDYDLQKALGPS-----HRTDCT-----DENPKLNRYC 1143
QY 1072 RSLPASEGSGSVDFG---DLGMAAGLQSLPHTDPSPLORYSEDTVPPLSETDGYV 1128
Db 1144 KPSNKNSSGDDERDSSAREVGVGNLR-----LDLPVDEDDYL 1182
QY 1129 APLTCSPOEVYVQPDVVRPQPPSPREGPLPAARAGATLERAKTLSPGKNGVVKVFAFG 1188
Db 1183 MP--TCQFGPNNNNMN-----NPNQNNMAVGAAGYM-----DLIGVP 1220
QY 1189 GAVENPEYL---TPQGGAAPOH-----PPPAFSP-AFONLYYWD 1224
Db 1221 VSDNPEYLLNAQTLGVGSEPIQTQIGIPVWGPGTMEVKVMPGSEPTSSDHEYND 1279

RESULT 14

S35745
protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus
C:Species: avian erythroblastosis virus
C>Date: 03-Mar-1994 #sequence_revision 26-May-1995 #text_change 28-Feb-1997
C:Accession: S35745
R:Vennstrom, B.
submitted to the EMBL Data Library, March 1993
A:Reference number: S35743
A:Accession: S35745
A:Molecule type: DNA
A:Residues: 1-544 <VEN>
A:Cross-references: EMBL:X12707
C:Genetics:
A:Gene: erbB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific P
F;135-400/Domain: protein kinase homology <KIN>
F;143-151/Region: protein kinase ATP-binding motif
F;170/Active site: Lys #status predicted

Query Match 24.2%; Score 1647; DB 2; Length 544;
Best Local Similarity 54.9%; Pred. No. 1.5e-60;
Matches 345; Conservative 70; Mismatches 121; Indels 92; Gaps 15;

QY 578 GPEADOCVACAHYKDPFCVACPSGVKPDLSYMPIWKFPDEGACQPCPINCTHSCVDL 637
Db 1 GP--DHCMKCAHFIDGPHCVKACPAAGVLGENDTL--VMKYADANAVCOLCHPNCTCKGCP 57

QY 638 DKGCPAEPQASPLTSIVSAVV--GILLVVVLGVVFGILIKRQOKIRKYTMRRLLQETEL 696
Db 58 GLEGCP---NGSKTFESIAAGVVGGLLCLVVVGLGILYLR--HIVRKTRLLRLLQEREL 113
QY 697 VPPLTPSGAMPNQAOIRILKETELRKVKYLGSGAGTVYKGIWIDGENVKIPVAIKVLR 756
Db 114 VPPLTPSGAMPNQAOIRILKETELRKVKYLGSGAGTVYKGIWIDGENVKIPVAIKVLR 173
QY 757 ENTPSKANKEIIDEAYVMAGVGSPPVSRILGICLTSTVOLVQMPYGLLDHVRNRCGR 816
Db 174 EATSPKANKEIIDEAYVMASVDNPHVCRLLGICLTSTVOLVQMPYGLLDVIREHKON 233
QY 817 LGSODLLANCMQIAKMSVLEDRVLVHRDLAARNVLVKSPPNHVKITDFGLARLLDIDETE 876
Db 234 IGSQVLLNLCVQIAKGMNLYLEERHVMHVRDLAARNVLVKTPOHVKITDFGLAKOLGADEKE 293
QY 877 YHADGGKVPKWMALSIILRRFTHQSDVWSGVTVWELMTGAKDYDGPAREIPDLLE 936
Db 294 YHAGGGKVPKWMALSIILHRYTHQSDVWSGVTVWELMTGSKDYDGPASEISVLE 353
QY 937 KGERLPQPPICITIDVYIMVIMVKMIDSECRPFRELVSEFSRMARDPQRFVVIQ--NEDLG 995
Db 354 KGERLPQPPICITIDVYIMVIMVKMSDADSRPKFRELIAEFSKWARDPPRYLVIOGDERMH 413
QY 996 PASPLDSTFYRSLLEDDMGDLVDAEYELVPOQGFPCPDPAAGAGVHHRSSSTRSG 1055
Db 414 LPSPTDSKYRTLMBEEDMEDIVDAEYLVPHQGF-----NSPST--- 454
QY 1056 GGDLTGLGLEPSEERAPSL-----APSEGAGSDVFDGLGMAAGLQSLPHTDPSPLQ 1110
Db 455 -----SRTPLLSLSATSNNSTATNCIDRNGG-----H----- 481
QY 1111 RYSEDPVPLPSETDGYVAPLTCSPQEVYVQPDVVRPQPPSPREGPLPAARAGAT--LER 1169
Db 482 -----PVREDGFL-----PAPEYVQ--LMPKPESTAMVQNYIYSLTAISK 523
QY 1170 AKTLSPGKNGVVKVFAFGGAVENPEYL 1197
Db 524 LPIDSRYQN-----SHSTAVDNPEYL 544

RESULT 15

S00727
kinase-related transforming protein (erbB) (EC 2.7.1.-) - avian erythroblastosis virus (1
C:Species: avian erythroblastosis virus
C>Date: 01-Dec-1989 #sequence_revision 01-Dec-1989 #text_change 28-Feb-1997
C:Accession: S00727
R:Scotting, P.; Vennstrom, B.; Jansen, M.; Graf, T.; Beug, H.; Hayman, M.J.
Oncogene Res. 1, 265-278, 1987
A:Title: Common site of mutation in the erbB gene of avian erythroblastosis virus mutant
A:Reference number: S00727; MUID:88217326; PMID:2897102
A:Accession: S00727
A:Molecule type: DNA
A:Residues: 1-545 <SCO>
A:Cross-references: EMBL:X06943
C:Genetics:
A:Gene: erbB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; phosphotransferase
F;135-400/Domain: protein kinase homology <KIN>
F;143-151/Region: protein kinase ATP-binding motif

Query Match 24.1%; Score 1640; DB 2; Length 545;
Best Local Similarity 54.9%; Pred. No. 3e-60;
Matches 345; Conservative 69; Mismatches 122; Indels 92; Gaps 15;

QY 578 GPEADOCVACAHYKDPFCVACPSGVKPDLSYMPIWKFPDEGACQPCPINCTHSCVDL 637
Db 1 GP--DHCMKCAHFIDGPHCVKACPAAGVLGENDTL--VMKYADANAVCOLCHPNCTCKGCP 57

QY 638 DKGCPAEPQASPLTSIVSAVV--GILLVVVLGVVFGILIKRQOKIRKYTMRRLLQETEL 696
Db 58 GLEGCP---NGSKTFESIAAGVVGGLLCLVVVGLGILYLR--HIVRKTRLLRLLQEREL 113

```

QY 697 VEPLTPSGAMPNOAIRKTELKVKVLGSGAFGTGVIWIPDGENVKIPVAIKVL 756
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 114 VEPLTPSGEAPNOAHLRIKTEBKKVKVLGFGAFGTGVIWIPGEKVTIPVAIKEL 173
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 757 ENTSPKANKIILDEAYVMAGVSPYVSRLLGICLTSTVQLVTOLMPYGCILLDHVRENRR 816
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 174 EATSPKANKIILDEAYVMASVDPHVCRLLGICLTSTVQLVTOLMPYGCILLDVIREHKN 233
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 817 LGSQDILLNWCQIAKNSYLEDLVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDE 876
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 234 IGSQYLLNWCQIAKNSYLEDLVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDE 293
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 877 YHADGGKVPKMALESILRRRTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLE 936
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 294 YHADGGKVPKMALESILHRIYTHQSDVMSYGVTVWELMTFGKPYDGIPIASEISSVLE 353
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 937 KGERLPQPPICITIDVYIMVIMVCMWIDSECRPRFRELVSFSESRMARDPQRFVVIQ-NEDLG 995
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 354 KGERLPQPPICITIDVYIMVIMVCMWIDSECRPRFRELVSFSESRMARDPQRFVVIQ-NEDLG 413
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 996 PASPLDSTFYRSILLEDDMGDLVDAEYLYPQOGFFCPDPAPGAGGMVHRRHSSSTRSG 1055
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 414 LPSFTDSKFYRTLMEEDMEDIVDAEYLYPHQGF-----NSPST--- 454
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1056 GGDLLTGLPESEEAAPRPL-----APSEGAGSDVFDGDLGMAAKGLQSLPTHDPSPLO 1110
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 455 -----SRTPLLSLSATSNNSATNCIDRNG-----H----- 481
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1111 RYSEDPTVPLPSETDGYVAPLTCSPQPEYVNPQDVRPQPPSPREGPLPAARPAAGAT-LER 1169
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 482 -----PVREDGFL-----PAPEYVNPQ-LMPKPSSTAMVQNOIYNYISLTAISK 523
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1170 AKTLSPGKGVKDVFAFGAVENPEYL 1197
Db      :|||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 524 LPMDSRYON-----SHSTAVDNPEYL 544

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:05:43 ; Search time 33.0519 Seconds
(without alignments)
1574.881 Million cell updates/sec

Title: US-09-806-703A-4
Perfect score: 6812
Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYGLDVPV 1255

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	6806	99.9	1255	1 ERB2 HUMAN	P04626 homo sapien
2	5994	88.0	1237	1 ERB2_RAT	P06494 rattus norv
3	5984.5	87.9	1254	1 ERB2_MESAU	Q60553 mesocricetu
4	3166	46.5	1210	1 EGFR HUMAN	P00533 homo sapien
5	3145	46.2	1210	1 EGFR MOUSE	Q01279 mus musculu
6	3003.5	44.1	1308	1 ERB4 HUMAN	Q15303 homo sapien
7	2984	43.8	1308	1 ERB4_RAT	P62956 rattus norv
8	2717.5	39.9	1167	1 XMRK_XIPMA	P13388 xiphophorus
9	2440.5	35.8	1342	1 ERB3 HUMAN	P21860 homo sapien
10	2369.5	34.8	1339	1 ERB3_RAT	Q62799 rattus norv
11	1974	29.0	1426	1 EGFR DROME	P04412 drosophila
12	1749.5	25.7	634	1 ERBB_ALV	P00534 avian leuko
13	1703	25.0	604	1 ERBB_AVIER	P00535 avian eryth
14	1630	23.9	540	1 ERBB_AVIEU	P11273 avian eryth
15	1624	23.8	703	1 EGFR CHICK	P13387 gallus gall
16	1302	19.1	1323	1 LT23 CABEL	P24348 caenorhabdi
17	1142.5	16.8	245	1 ERB2 MOUSE	P70424 mus musculu
18	735	10.8	1363	1 ILPR BRALA	O02466 branchiost
19	716	10.5	1382	1 INSR HUMAN	P06213 homo sapien
20	711	10.4	1383	1 INSR_RAT	P15127 rattus norv
21	710.5	10.4	1372	1 INSR MOUSE	P15208 mus musculu
22	707	10.4	1300	1 IRR MOUSE	Q9wt14 mus musculu
23	698	10.2	1297	1 IRR_HUMAN	P14616 homo sapien
24	698	10.2	1607	1 MIPR LYVMT	Q25410 lymnaea sta
25	693.5	10.2	1300	1 IRR_CAVPO	P14617 cavia porce
26	685	10.1	1477	1 HTK_HYDAT	Q25197 hydra atten
27	652	9.6	1367	1 IGIR HUMAN	P08069 homo sapien
28	642	9.4	1373	1 IGIR MOUSE	P060751 mus musculu
29	638.5	9.4	1370	1 IG1R_RAT	P24062 rattus norv
30	626	9.2	1390	1 INSR_AEDAE	Q93105 aedes aegyp
31	619	9.1	2146	1 INSR_DROME	P09208 drosophila
32	606	8.9	987	1 EPB4 HUMAN	P54760 homo sapien
33	591.5	8.7	984	1 EPB1_CHICK	Q07494 gallus gall

34	589.5	8.7	977	1	EPA2 MOUSE	Q03145 mus musculu
35	588	8.6	1114	1	REP2 HUMAN	P07949 homo sapien
36	584.5	8.6	976	1	EPA2 HUMAN	P29317 homo sapien
37	584	8.6	987	1	EPB4 MOUSE	P54761 mus musculu
38	583.5	8.6	984	1	EPB1_RAT	P09759 rattus norv
39	579	8.5	902	1	EPBB_XENLA	Q91736 xenopus lae
40	577.5	8.5	984	1	EPB1_HUMAN	P54762 homo sapien
41	576.5	8.5	985	1	EPBA_XENLA	Q91571 xenopus lae
42	573.5	8.4	1053	1	FAK1 CHICK	Q00944 gallus gall
43	569	8.4	1068	1	FAK1_XENLA	Q91738 xenopus lae
44	567	8.3	757	1	HT16_HYDAT	P53356 hydra atten
45	563	8.3	1052	1	FAK1_MOUSE	P34152 mus musculu

ALIGNMENTS

RESULT 1						
ERB2_HUMAN	ID	ERB2_HUMAN	STANDARD;	PRT;	1255	AA.
AC	P04626;					
DT	13-AUG-1987	(Rel. 05, Created)				
DT	13-AUG-1987	(Rel. 05, Last sequence update)				
DT	15-JUN-2002	(Rel. 41, Last annotation update)				
DE	Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)					
DE	(p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Tyrosine kinase-type cell					
DE	surface receptor HER2) (MLN 19).					
GN	ERBB2 OR HER2 OR NGL OR NEU.					
OS	Homo sapiens (Human)					
OC	Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
OX	NCBI_TaxID=9606;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RX	MEDLINE=86118663; PubMed=3003577;					
RA	Yamamoto T., Ikawa S., Akiyama T., Semba K., Nomura N., Miyajima N.,					
RA	Saito T., Toyoshima K.;					
RT	"Similarity of protein encoded by the human c-erb-B-2 gene to					
RT	epidermal growth factor receptor.;"					
RL	Nature 319:230-234(1986).					
RN	[2]					
RP	SEQUENCE FROM N.A.					
RX	MEDLINE=86070181; PubMed=2999974;					
RA	Cousens L., Yang-Feng T.L., Liao Y.C., Chen E., Gray A.,					
RA	McGrath J., Seeburg P.H., Libermann T.A., Schlessinger J.,					
RA	Francke U., Levinson A., Ullrich A.;					
RT	"Tyrosine kinase receptor with extensive homology to EGF receptor					
RT	shares chromosomal location with neu oncogene.;"					
RL	Science 230:1132-1139(1985).					
RN	[3]					
RP	SEQUENCE OF 737-1031 FROM N.A.					
RX	MEDLINE=86016729; PubMed=2995967;					
RA	Semba K., Kamata N., Toyoshima K., Yamamoto T.;					
RT	"A v-erbB-related protooncogene, c-erbB-2, is distinct from the					
RT	c-erbB-1/epidermal growth factor-receptor gene and is amplified in a					
RT	human salivary gland adenocarcinoma.;"					
RL	Proc. Natl. Acad. Sci. U.S.A. 82:6497-6501(1985).					
RN	[4]					
RP	VARIANTS VAL-654 AND VAL-655.					
RX	MEDLINE=93194196; PubMed=8095488;					
RA	Ehsani A., Low J., Wallace R.B., Wu A.M.;					
RT	"Characterization of a new allele of the human ERBB2 gene by allele-					
RT	specific competition hybridization.;"					
RL	Genomics 15:426-429(1993).					
CC	-!- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,					
CC	ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A					
CC	POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-					
CC	ALPHA AND AMPHIREGULIN.					
CC	-!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein					
CC	tyrosine phosphate.					
CC	-!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS					
CC	(POTENTIAL).					
CC	-!- SUBCELLULAR LOCATION: Type I membrane protein.					

-1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
RESIDUES (BY SIMILARITY).
-1- POLYMORPHISM: THERE ARE FOUR ALLELES DUE TO THE VARIATIONS IN
POSITIONS 654 AND 655. ALLELE B1 (654-ILE-ILE-655) HAS A FREQUENCY
OF 0.782; ALLELE B2 (654-ILE-VAL-655) HAS A FREQUENCY OF 0.206;
ALLELE B3 (654-VAL-VAL-655) HAS A FREQUENCY OF 0.012.
-1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; M11767; AAA35808.1; -
DR EMBL; M11767; AAA35808.1; JOINED.
DR EMBL; M11762; AAA35808.1; JOINED.
DR EMBL; M11763; AAA35808.1; JOINED.
DR EMBL; M11764; AAA35808.1; JOINED.
DR EMBL; M11765; AAA35808.1; JOINED.
DR EMBL; M11766; AAA35808.1; JOINED.
DR EMBL; M11767; AAA35808.1; JOINED.
DR EMBL; M12036; AAA35978.1; -
DR EMBL; M12036; AAA35978.1; -
DR EMBL; X03363; CAA27060.1; -
DR PIR; A25491; A25491.
DR PIR; A24571; A24571.
DR HSP; P11362; IrgK.
DR HSP; H03363; CAA27060.1; -
DR MIM; 164870; -
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR000719; Euk_Pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR InterPro; IPR004019; YLP_motif.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L domain; 2.
DR Pfam; PF02757; YLP_2 domain; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00109; PROTEIN KINASE_TYR; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW Polymorphism.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 1255 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.
FT DOMAIN 22 652 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 653 675 POTENTIAL.
FT DOMAIN 676 1255 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 720 987 PROTEIN KINASE.
FT NP_BIND 726 734 ATP (BY SIMILARITY).
FT BINDING 753 753 ATP (BY SIMILARITY).
FT ACT_SITE 845 845 BY SIMILARITY.
FT DISULFID 195 204 BY SIMILARITY.
FT DISULFID 199 212 BY SIMILARITY.
FT DISULFID 220 227 BY SIMILARITY.
FT DISULFID 224 235 BY SIMILARITY.
FT DISULFID 236 244 BY SIMILARITY.
FT DISULFID 240 252 BY SIMILARITY.
FT DISULFID 255 264 BY SIMILARITY.
FT DISULFID 268 295 BY SIMILARITY.
FT DISULFID 299 311 BY SIMILARITY.
FT DISULFID 315 331 BY SIMILARITY.
FT DISULFID 334 338 BY SIMILARITY.
FT DISULFID 511 520 BY SIMILARITY.
FT DISULFID 515 528 BY SIMILARITY.
FT DISULFID 531 540 BY SIMILARITY.
FT DISULFID 544 560 BY SIMILARITY.

FT	DISULFID	563	576	BY SIMILARITY.
FT	DISULFID	567	584	BY SIMILARITY.
FT	DISULFID	587	596	BY SIMILARITY.
FT	DISULFID	600	623	BY SIMILARITY.
FT	DISULFID	626	634	BY SIMILARITY.
FT	DISULFID	630	642	BY SIMILARITY.
FT	MOD_RES	1139	1139	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	MOD_RES	1248	1248	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	CARBOHYD	68	68	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	124	124	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	187	187	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	259	259	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	530	530	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	571	571	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	629	629	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	VARIANT	654	654	I -> V.
FT	VARIANT	655	655	/FTID=VAR_004077.
FT	CONFLICT	1170	1170	/FTID=VAR_004078.
FT	SEQUENCE	1255	AA; 137909	MM; 39E9DFA04DCF962 CRC64;

Query Match 99.9%; Score 6806; DB 1; Length 1255;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1253; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY	1	MELALCRWGLLLALLPPGAAS	TOVCTG	TDKMLRLPAS	PETHLDMLRLHYQGC	VOVQGNL	60
DB	1	MELALCRWGLLLALLPPGAAS	TOVCTG	TDKMLRLPAS	PETHLDMLRLHYQGC	VOVQGNL	60
QY	61	ELTYLPTNASLSFLQDI	QEVQGV	VLIAHQVQVPLQRL	RIVRGTQ	LFEDNYALAV	120
DB	61	ELTYLPTNASLSFLQDI	QEVQGV	VLIAHQVQVPLQRL	RIVRGTQ	LFEDNYALAV	120
QY	121	DPLNNTPTVTGASPGGL	RELQRL	SLTEILKGGVLI	QORNPOLCYQD	TILWKDIF	180
DB	121	DPLNNTPTVTGASPGGL	RELQRL	SLTEILKGGVLI	QORNPOLCYQD	TILWKDIF	180
QY	181	LTLIDNRSRACHPCSP	CKGRCW	GESSEDCQSL	TRTYCAGC	CKRCKPLPTD	240
DB	181	LTLIDNRSRACHPCSP	CKGRCW	GESSEDCQSL	TRTYCAGC	CKRCKPLPTD	240
QY	241	AAGCTGPKHSDCLAC	LHFHNSG	ICELHCPALV	TYNTDTFES	MPNPEGRYTF	300
DB	241	AAGCTGPKHSDCLAC	LHFHNSG	ICELHCPALV	TYNTDTFES	MPNPEGRYTF	300
QY	301	YNYLSTDVGSCTLV	CPHMQE	VTAEADGTOR	CKSKPCAR	VCYGLGMEHL	360
DB	301	YNYLSTDVGSCTLV	CPHMQE	VTAEADGTOR	CKSKPCAR	VCYGLGMEHL	360
QY	361	IOEFAGCKKIFGSLA	FLPES	FDGDPASNTA	PLQPEOLQ	VFETLEET	420
DB	361	IOEFAGCKKIFGSLA	FLPES	FDGDPASNTA	PLQPEOLQ	VFETLEET	420
QY	421	DLISVFONLQVIRGR	ILHNGAYS	SLTQGLG	ISWGLRSL	RELSGSLAL	480
DB	421	DLISVFONLQVIRGR	ILHNGAYS	SLTQGLG	ISWGLRSL	RELSGSLAL	480
QY	481	PWDQLFRNPHQALL	HTANR	PEDEC	CVGEGLA	CHQLCARGH	540
DB	481	PWDQLFRNPHQALL	HTANR	PEDEC	CVGEGLA	CHQLCARGH	540
QY	541	VEECRVLOGLPREY	VNARHCL	PCHEP	QONSGV	TCFGEAD	600
DB	541	VEECRVLOGLPREY	VNARHCL	PCHEP	QONSGV	TCFGEAD	600
QY	601	PSGVKPDLSYMPI	WKFP	PPDEGAC	QPCP	INCTHSC	660
DB	601	PSGVKPDLSYMPI	WKFP	PPDEGAC	QPCP	INCTHSC	660
QY	661	ILLVWVVGVL	FGILIKR	ROOKIRK	YTMRL	LQETEL	720
DB	661	ILLVWVVGVL	FGILIKR	ROOKIRK	YTMRL	LQETEL	720

QY 721 RKVKVLGSGAGFTYKGIWIPDGENVKIPVAIKVLRENTSPKANKIILDYAYMAGVGSP 780
 DB 721 RKVKVLGSGAGFTYKGIWIPDGENVKIPVAIKVLRENTSPKANKIILDYAYMAGVGSP 780
 QY 781 YVSRLLGICLTSTVOLVTLQMPYGLLDHVRNRLGSDLLNCWCQIAKGSYLEDVR 840
 DB 781 YVSRLLGICLTSTVOLVTLQMPYGLLDHVRNRLGSDLLNCWCQIAKGSYLEDVR 840
 QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPVKWMALESILRRFT 900
 DB 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPVKWMALESILRRFT 900
 QY 901 HQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPPOPICTIDVYIMVKCWM 960
 DB 901 HQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPPOPICTIDVYIMVKCWM 960
 QY 961 IDSECRPRFRELSESRMARDPQRFVIONEDLGPASPLDSTFYRSLLEDMDGLVDA 1020
 DB 961 IDSECRPRFRELSESRMARDPQRFVIONEDLGPASPLDSTFYRSLLEDMDGLVDA 1020
 QY 1021 EYLVVQOQFCFDPAPAGAGVHHRSSSTRSGGDLTLGLEPSEEAAPRSLAPSEG 1080
 DB 1021 EYLVVQOQFCFDPAPAGAGVHHRSSSTRSGGDLTLGLEPSEEAAPRSLAPSEG 1080
 QY 1081 AGSDVFDGDLGMAAGKQLSLTHDPSPLQRYSEDPTVPLPSETDGVAPLTCSPQPEYV 1140
 DB 1081 AGSDVFDGDLGMAAGKQLSLTHDPSPLQRYSEDPTVPLPSETDGVAPLTCSPQPEYV 1140
 QY 1141 NQPDVPPPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAGGAVENPEYLTPO 1200
 DB 1141 NQPDVPPPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAGGAVENPEYLTPO 1200
 QY 1201 GGAAPQHPHPPAFSPAFDNLVYWDQPPPERGAPSTFKGPTAENPEYLGLDVPEV 1255
 DB 1201 GGAAPQHPHPPAFSPAFDNLVYWDQPPPERGAPSTFKGPTAENPEYLGLDVPEV 1255

RESULT 2
 ID_ERB2_RAT STANDARD; PRT; 1257 AA.
 AC P06494;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
 DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Epidermal growth factor
 DE receptor-related protein).
 DE ERBB2 OR NEU.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RC TISSUE=FROM N.A.
 RC TISSUE=Neuroblastoma;
 RX MEDLINE=86118662; PubMed=3945311;
 RA Bargmann C.I., Hung M.-C., Weinberg R.A.;
 RT "The neu oncogene encodes an epidermal growth factor receptor-related
 RT protein.";
 RL Nature 319:226-230 (1986).
 RN [2]
 RP SEQUENCE OF 852-905 FROM N.A.
 RC TISSUE=Sciatic nerve;
 RX MEDLINE=91222560; PubMed=2025425;
 RA Lai C., Lemke G.;
 RT "An extended family of protein-tyrosine kinase genes differentially
 RT expressed in the vertebrate nervous system.";
 RN Neuron 6:691-704 (1991).
 RN [3]
 RP STRUCTURE BY NMR OF 650-668.
 RX MEDLINE=92155181; PubMed=1346763;
 RA Gullick W.J., Bottomley A.C., Lofts F.J., Doak D.G., Mulvey D.,

RA Newman R., Crumpton M.J., Sternberg M.J.E., Campbell I.D.;
 RT "Three dimensional structure of the transmembrane region of the proto-
 RT oncogenic and oncogenic forms of the neu protein.";
 RL EMBO J. 11:43-48 (1992).
 CC -!- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
 CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
 CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
 CC ALPHA AND AMPHIREGULIN.
 CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS.
 CC THE CONSTITUTIVELY ACTIVATED ONCOGENIC VARIANT FORMS A HOMODIMER.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
 CC RESIDUES (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X03362; CAA27059.1; ALT_INIT.
 DR PIR; A24562; TVRTNU.
 DR HSSP; P11362; LFQG.
 DR InterPro; IPR000494; EGFR_L_domain.
 DR InterPro; IPR000719; Euk_pkinase.
 DR InterPro; IPR002174; Furin-like.
 DR InterPro; IPR001245; Tyr_pkinase.
 DR InterPro; IPR004019; YLP motif.
 DR Pfam; PF00069; pkinase; 1.
 DR Pfam; PF00757; Furin-like; 1.
 DR Pfam; PF01030; Recep_L_domain; 2.
 DR Pfam; PF02757; YLP; 2.
 DR ProDom; PD000001; Euk_pkinase; 1.
 DR SMART; SM00261; FU; 3.
 DR SMART; SM00219; TyrKC; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE; PS0011; PROTEIN_KINASE_DOM; 1.
 KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
 KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
 KW Proto-oncogene; Disease mutation.
 FT SIGNAL 1 21 POTENTIAL.
 FT CHAIN 22 1257 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.
 FT DOMAIN 22 654 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 655 677 POTENTIAL.
 FT DOMAIN 678 1257 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 159 369 CYS-RICH.
 FT DOMAIN 473 646 CYS-RICH.
 FT DOMAIN 722 989 PROTEIN KINASE.
 FT NP BIND 728 736 ATP (BY SIMILARITY).
 FT BINDING 755 755 ATP (BY SIMILARITY).
 FT ACT_SITE 847 847 BY SIMILARITY.
 FT DISULFID 196 205 BY SIMILARITY.
 FT DISULFID 200 213 BY SIMILARITY.
 FT DISULFID 221 228 BY SIMILARITY.
 FT DISULFID 225 236 BY SIMILARITY.
 FT DISULFID 237 245 BY SIMILARITY.
 FT DISULFID 241 253 BY SIMILARITY.
 FT DISULFID 256 265 BY SIMILARITY.
 FT DISULFID 269 296 BY SIMILARITY.
 FT DISULFID 300 312 BY SIMILARITY.
 FT DISULFID 316 332 BY SIMILARITY.
 FT DISULFID 335 339 BY SIMILARITY.
 FT DISULFID 513 522 BY SIMILARITY.
 FT DISULFID 517 530 BY SIMILARITY.
 FT DISULFID 533 542 BY SIMILARITY.
 FT DISULFID 546 562 BY SIMILARITY.
 FT DISULFID 565 578 BY SIMILARITY.

FT	DISULFID	569	586	BY SIMILARITY.
FT	DISULFID	589	598	BY SIMILARITY.
FT	DISULFID	602	625	BY SIMILARITY.
FT	DISULFID	628	636	BY SIMILARITY.
FT	DISULFID	632	644	BY SIMILARITY.
FT	MOD_RES	1141	1141	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	MOD_RES	1250	1250	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	CARBOHYD	68	68	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	188	188	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	260	260	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	532	532	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	573	573	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	631	631	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARIANT	661	661	N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ	SEQUENCE	1257 AA; 138831 MW; 6129264583011402 CRC64;		V -> E (IN ONCOGENIC NEU).
Query Match				
Best Local Similarity				
Matches 1104; Conservative 50; Mismatches 101; Indels 2; Gaps 2;				
Qy	1	MELAAALCRWGLLLALLPPGAASTQVCTGDMKRLRPASPETHLDMLRHLYGCCVVQGNL	60	
Db	1	MELAAACRWGFLALLPPGIAGTQVCTGDMKRLRPASPETHLDMLRHLYGCCVVQGNL	60	
Qy	61	ELTYLPNASLSFLQDIQEOVGYYVLIIAHNQVRQVPLQRLRIVRGTFQLPEDNYALAVLDNG	120	
Db	61	ELTYVPANASLSFLQDIQEOVGYYMLIAHNQVRVPLQRLRIVRGTFQLPEDKYALAVLDNR	120	
Qy	121	DLPLNNTTPVT- GASPGGIRELOLRSLTEILKGGVLIQBNPOLCYODTTLWKDIAPHKNQL	179	
Db	121	DPQDNVAASTPGRTPGELRGELRLQSLTEILKGGVLIIRGNPOLCYQDMVLWKDVFRRKNQL	180	
Qy	180	ALTLLIDNTRSRACHPCSPMKGSGCWGESSBDCSLRTVCAGGCAKCKGRLPTDCCHEQ	239	
Db	181	APVDIDNTRSPACPPCAPACKDNHCWGESPEDCQILGTICTSGCAKCKGRLPTDCCHEQ	240	
Qy	240	CAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMNPPEGRYTFGASCVTAC	299	
Db	241	CAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMNPPEGRYTFGASCVTTC	300	
Qy	300	PYNVLSIDVGSCTLVCLPHNQEVTAEDGTQRCCKSPCARVCYGLGMEHLREVRVATSA	359	
Db	301	PYNVLSIEVSGSCTLVCPFPNNQEVTAEDGTQRCCKSPCARVCYGLGMEHLRGARAITSD	360	
Qy	360	NIQEFACCKIFGSLAFLPESFDGDPASNTAPLOEQVPELTLEEITGYLYISAWPDSL	419	
Db	361	NVEQFDCKIFGSLAFLPESFDGDPSSGTAPLRPEQLQVPELTLEEITGYLYISAWPDSL	420	
Qy	420	PDLSVFONLQVIRGRIILHNGAYSITLQGLGISWLGSLRSLRGSLALIHHTHLCFVHT	479	
Db	421	RDLVSFONLRIIRGRIILDGAYSITLQGLGHSILGSLRSLRGSLALIHNAHLCFVHT	480	
Qy	480	VPMDQLFRNPHOALLHTANRPEDE- CVGEGLAACHQLCARGHCWGPGPTQCNCQFIRGQ	538	
Db	481	VPMDQLFRNPHOALLHSGNRPEEDLVCSSGLVCNLSLCAHGCWCGPGPTQCNCSHFLRGQ	540	
Qy	539	ECVEECRVLOGLPREYVYNARHCLPCHPECOFGNSVTCFCGPEADQCVACAHVKDPPFCVA	598	
Db	541	ECVEECKWKLGPREYVSDKRLPCHPECOFQNSSETCFGSEADQCAACAHYKDDSSSCVA	600	
Qy	599	RCPSGVKPDLSYMPIMKFPDPEEGACQPCPINCTHSCVDLDKGCPAEORASPLTISYSAV	658	
Db	601	RCPSGVKPDLSYMPIMKYDDEEGICQPCPINCTHSCVDLDERGCPAEORASPVTFIATV	660	
Qy	659	VGILLVVVLGVVFGILLIKRQOKIRKYTMRLLOQETELVEPLTPSGAMPNQAQMRILKET	718	
Db	661	VGVLFLILVVVGLIKRKRQKIRKYTMRLLOQETELVEPLTPSGAMPNQAQMRILKET	720	
Qy	719	ELRKVKVLGSGAFCTVYKGWIIPDGENVKIPVALIKVIRENTSPKANKEILDEAYVMAGV	778	
Db	721	ELRKVKVLGSGAFCTVYKGWIIPDGENVKIPVALIKVIRENTSPKANKEILDEAYVMAGV	780	
Qy	779	SPVYSRLIGLCTSTVOLVTQLPYGCLLDHRNRRNRLGSLQDILLNWCQITAGMSYLED	838	

entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC	EMBL; D16295; BAA03801.1; -	QY	61	ELTYLPTNASLSFLQDIOEVQGYVLIHANQVQVFLQRLRIVRGTLQFEDNYVALAVLDNG	120
CC	HSSP; P11362; LFQK.	Db	61	ELTYLPANATLSFLQDIOEVQGYVLIHANQVQVFLQRLRIVRGTLQFEDNYVALAVLDNR	120
DR	InterPro; IPR000494; EGFR_L domain.	QY	121	DPLNNTPTVGTASPGGLRELQRLSITELKGGVLTORNPOLCYODTILWKDIFHKNQOLA	180
DR	InterPro; IPR000719; Euk_pkinase.	Db	121	DPLDNTVTATGRTPEGLRELQRLSITELKGGVLTORNPOLCYODTILWKDIFHKNQOLA	180
DR	InterPro; IPR002174; Furin-like.	QY	181	LTLIDNRSRACHPCSPMCKSGRCWGESSEDCQSILTRTVCAGGCARCKGKPLPTDCCHEQC	240
DR	InterPro; IPR004019; YLP motif.	Db	181	PVDIDNRSRACHPCSPMCKSGRCWGESSEDCQSILTRTVCAGGCARCKGKPLPTDCCHEQC	240
DR	pfam; PF00757; Furin-like; 1.	QY	241	AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRVTFGASCVTACP	300
DR	pfam; PF00757; Furin-like; 1.	Db	241	AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRVTFGASCVTTC	300
DR	pfam; PF01030; Recep_L domain; 2.	QY	301	YNYLSTDVGSCTLVCPLNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVATSAN	360
DR	pfam; PF01030; Recep_L domain; 2.	Db	301	YNYLSTDVGSCTLVCPLNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVATSAN	360
DR	SMART; SM00219; TyrKC; 1.	QY	361	IOEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPELOQVLFETLEETGYLYISAWPDSL	420
DR	PROSITE; PS00107; PROTEIN KINASE ATP; 1.	Db	361	IOEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPELOQVLFETLEETGYLYISAWPDSL	420
DR	PROSITE; PS00109; PROTEIN KINASE TTR; 1.	QY	421	DLVSFQNLQVIRGRILHNGAYSILTQGLGSIWGLRLSRLSRELGLALIHNNHLCFVHTV	480
DR	PROSITE; PS50011; PROTEIN KINASE DOM; 1.	Db	421	DLVSFQNLQVIRGRILHNGAYSILTQGLGSIWGLRLSRLSRELGLALIHNNHLCFVHTV	480
KW	Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;	QY	481	PWDQLFRNPHQALLHTANRPEDECYVGEGLACHQLCARGHCWGPPTQCVCNCSQFLRGQEC	540
KW	Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;	Db	481	PWDQLFRNPHQALLHTANRPEDECYVGEGLACHQLCARGHCWGPPTQCVCNCSQFLRGQEC	540
KW	Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;	QY	541	VEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFCGPEADQCVACAHVKDPPPCVARC	600
KW	Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;	Db	541	VEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFCGPEADQCVACAHVKDPPPCVARC	600
FT	Proto-oncogene; Disease mutation.	QY	601	PSGVKPDLSYMPIMKFPDEEGACQPCINCHTHSCVDLDDKGCPCAEQASPLTSIVSAVVG	660
FT	SIGNAL	Db	601	PSGVKPDLSYMPIMKFPDEEGACQPCINCHTHSCVDLDDKGCPCAEQASPLTSIVSAVVG	660
FT	CHAIN	QY	661	ILLVVLGVVFGILIKRROQKIRKVTMRLLQETELVEPLTPSGAMPNQAORILKTEL	720
FT	DOMAIN	Db	661	ILLVVLGVVFGILIKRROQKIRKVTMRLLQETELVEPLTPSGAMPNQAORILKTEL	720
FT	NP_BIND	QY	721	RKVKVLSGAFCTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMAGVSP	780
FT	BINDING	Db	721	RKVKVLSGAFCTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMAGVSP	780
FT	ACT_SITE	QY	781	YVSRLLGICLTSTVOLVTQMPYGCILDHVRENRGLSGQDILLNMCQIAKMSYLEYVR	840
FT	DISULFID	Db	781	YVSRLLGICLTSTVOLVTQMPYGCILDHVRENRGLSGQDILLNMCQIAKMSYLEYVR	840
FT	DISULFID	QY	841	LVHRDLAARNVLKSPNVHKITDFGLARLLDIDETEHADGKVKPIKMALESILRRFT	900
FT	DISULFID	Db	841	LVHRDLAARNVLKSPNVHKITDFGLARLLDIDETEHADGKVKPIKMALESILRRFT	900
FT	DISULFID	QY	901	HOSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPCTIDVYMWVKWM	960
FT	DISULFID	Db	901	HOSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPCTIDVYMWVKWM	960
FT	DISULFID	QY	961	IDSECRPRPRELVSFSSRMARDPQRVFIQNEBGLPASPLDSTFYRSLLEDMDGLDVA	1020
FT	DISULFID	Db	961	IDSECRPRPRELVSFSSRMARDPQRVFIQNEBGLPASPLDSTFYRSLLEDMDGLDVA	1020
FT	DISULFID	QY	1021	ESYLVQQGFFCPDPAPAGAGMWHHRSSSTRSGGGLTLGLEPSEEPASPLAPSG	1080
FT	DISULFID	Db	1021	ESYLVQQGFFCPDPAPAGAGMWHHRSSSTRSGGGLTLGLEPSEEPASPLAPSG	1080
FT	DISULFID	QY	1081	AGSDVFDGDLGNAKGLQSLTPHDPSPLORYSEDPTVPLPSETDGYVAPLTCSPQEVV	1140
FT	DISULFID	Db	1081	AGSDVFDGDLGNAKGLQSLTPHDPSPLORYSEDPTVPLPSETDGYVAPLTCSPQEVV	1140

Query Match 87.9%; Score 5984.5; DB 1; Length 1254;
 Best Local Similarity 87.6%; Pred. No. 6.3e-311;
 Matches 1099; Conservative 58; Mismatches 97; Indels 1; Gaps 1;
 1 MELAAACRWGLLLALLPPGNASTQVCTGTDMLKRLPASPTHLDMLRHLYQGCVVQGNL 60
 1 MELAAACRWGLLLALLSPGASGTQVCTGTDMLKRLPASPTHLDIVRHLYQGCVVQGNL 60

Qy 1141 NQDVVRQPPSPREGPLPAARAGATLERAKTILSPKNGVVKVDFAFGAVENPEYLTPQ 1200
 Db 1141 NQDVVRQPPSPREGPLPAARAGATLERAKTILSPKNGVVKVDFAFGAVENPEYLTPQ 1200
 Qy 1201 GGAAPQHPAPSPAFDNLVYWDQDPPERGAPSTPKGTPTAENPPEYGLDVPV 1255
 Db 1201 GGSASQHP-PPALCPAFDNLVYWDQDPPSERGSPNPTPEGTPTAENPPEYGLDVPV 1254

RESULT 4

EGFR_HUMAN STANDARD; PRT; 1210 AA.

AC P00533; P06268; Q14225; Q9UMD7; Q9UMD8; Q9UMG5; Q92795; Q00732;
 AC O0688; Q9BZS2; Q9H2C9; Q9GZX1; Q9H3C9;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (Receptor
 DE protein-tyrosine kinase ErbB-1).
 GN EGFR OR ERBB1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RX MEDLINE=84219729; PubMed=6328312;
 RA Ullrich A., Coussens L., Hayflick J.S., Dull T.J., Gray A., Tam A.W.,
 RA Lee J., Yarden Y., Libermann T.A., Schlessinger J., Downward J.,
 RA Mayes E.L.V., Whittle N., Waterfield M.D., Seeburg P.H.;
 RT "Human epidermal growth factor receptor cDNA sequence and aberrant
 RT expression of the amplified gene in A431 epidermoid carcinoma cells.";
 RL Nature 309:418-425(1984).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RX TISSUE=Placenta;
 RC MEDLINE=95382957; PubMed=7654368;
 RA Ilekis J.V., Stark B.C., Scoccia B.;
 RT "Possible role of variant RNA transcripts in the regulation of
 RT epidermal growth factor receptor expression in human placenta.";
 RL Mol. Reprod. Dev. 41:149-156(1995).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RX TISSUE=Placenta;
 RC MEDLINE=97078686; PubMed=8918811;
 RA Reiter J.L., Mailhe N.J.;
 RT "A 1.8 kb alternative transcript from the human epidermal growth
 RT factor receptor gene encodes a truncated form of the receptor.";
 RL Nucleic Acids Res. 24:4050-4056(1996).
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RX TISSUE=Placenta;
 RC MEDLINE=97256547; PubMed=9103388;
 RA Ilekis J.V., Gariti J., Niederberger C., Scoccia B.;
 RT "Expression of a truncated epidermal growth factor receptor-like
 RT protein (TEGFR) in ovarian cancer.";
 RL Gynecol. Oncol. 65:36-41(1997).
 RN [5]
 RP SEQUENCE FROM N.A. (ISOFORMS 3 AND 4).
 RX TISSUE=Placenta;
 RC MEDLINE=21100872; PubMed=11161793;
 RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
 RA Schehl Sinclair C., Peareall R.S., Green P.J., Yee D., Lampland A.L.,
 RA Balasubramanian S., Crossley T.R., Magnuson T.R.,
 RA Mailhe N.J.;
 RT "Comparative genomic sequence analysis and isolation of human and
 RT mouse alternative EGFR transcripts encoding truncated receptor
 RT isoforms.";
 RL Genomics 71:1-20(2001).
 RN [6]
 RP SEQUENCE OF 575-687 FROM N.A.
 RA Reiter J.L., Threadgill D.W., Danielsen A.J., Schehl C.M.,
 RA Lampland A.L., Balasubramanian S., Crossley T.O., Magnuson T.R.,

RA Mailhe N.J.;
 RT "Human and mouse alternative EGFR transcripts encoding only the
 RT extracellular domain of the receptor";
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 RN [7]
 RP SEQUENCE OF 713-924 FROM N.A.
 RX MEDLINE=84196372; PubMed=6326261;
 RA Lin C.R., Chen W.S., Krueger W., Stolarsky L.S., Weber W.,
 RA Evans R.M., Verma I.M., Gill G.N., Rosenfeld M.G.;
 RT "Expression cloning of human EGF receptor complementary DNA: gene
 RT amplification and three related messenger RNA products in A431
 RT cells.";
 RL Science 224:843-848(1984).
 RN [8]
 RP SEQUENCE OF 150-962 FROM N.A.
 RX MEDLINE=84245835; PubMed=6330563;
 RA Xu Y.H., Ithil S., Clark A.J.L., Sullivan M., Wilson R.K., Ma D.P.,
 RA Roe B.A., Merlino G.T., Pastan I.;
 RT "Human epidermal growth factor receptor cDNA is homologous to a
 RT variety of RNAs overproduced in A431 carcinoma cells.";
 RL Nature 309:806-810(1984).
 RN [9]
 RP SEQUENCE OF 1028-1210 FROM N.A.
 RX MEDLINE=85046483; PubMed=6093780;
 RA Simmen F.A., Gope M.L., Schulz T.Z., Wright D.A., Carpenter G.,
 RA O'Malley B.W.;
 RT "Isolation of an evolutionarily conserved epidermal growth factor
 RT receptor cDNA from human A431 carcinoma cells.";
 RL Biochem. Biophys. Res. Commun. 124:125-132(1984).
 RN [10]
 RP SEQUENCE OF 1-29 FROM N.A.
 RX MEDLINE=88217333; PubMed=3329716;
 RA Haley J.D., Whittle N., Bennett P., Kinchington D., Ullrich A.,
 RA Waterfield M.D.;
 RT "The human EGF receptor gene: structure of the 110 kb locus and
 RT identification of sequences regulating its transcription.";
 RL Oncogene Res. 1:375-396(1987).
 RN [11]
 RP SEQUENCE OF 1-29 FROM N.A.
 RX MEDLINE=91107677; PubMed=1988448;
 RA Haley J.D., Waterfield M.D.;
 RT "Contributory effects of de novo transcription and premature
 RT transcript termination in the regulation of human epidermal growth
 RT factor receptor proto-oncogene RNA synthesis.";
 RL J. Biol. Chem. 266:1746-1753(1991).
 RN [12]
 RP SEQUENCE OF 1-29 FROM N.A.
 RX MEDLINE=85270438; PubMed=2991899;
 RA Ishii S., Xu Y.H., Stratton R.H., Roe B.A., Merlino G.T., Pastan I.;
 RT "Characterization and sequence of the promoter region of the human
 RT epidermal growth factor receptor gene.";
 RL Proc. Natl. Acad. Sci. U.S.A. 82:4920-4924(1985).
 RN [13]
 RP SEQUENCE OF 540.
 RA Kohda D.;
 RL Submitted (SEP-1997) to the SWISS-PROT data bank.
 RN [14]
 RP RECEPTOR ACTIVITY.
 RX MEDLINE=84191554; PubMed=6325948;
 RA Mroczkowski B., Mosig G., Cohen S.;
 RT "ATP-stimulated interaction between epidermal growth factor receptor
 RT and supercoiled DNA.";
 RL Nature 309:270-273(1984).
 RN [15]
 RP PHOSPHORYLATION.
 RX MEDLINE=89278137; PubMed=2543678;
 RA Margolis B.L., Lax I., Kris R., Dombalagian M., Honegger A.M.,
 RA Hawk R., Givol D., Ullrich A., Schlessinger J.;
 RT "All autophosphorylation sites of epidermal growth factor (EGF)
 RT receptor and HER2/neu are located in their carboxyl-terminal tails.
 RT Identification of a novel site in EGF receptor.";
 RL J. Biol. Chem. 264:10667-10671(1989).
 RN [16]

RP CARBOHYDRATE-LINKAGE SITES ASN-128; ASN-175; ASN-413; ASN-444 AND
 RP ASN-528.
 RX MEDLINE=96398132; PubMed=8962717;
 RA Smith K.D., Davies M.J., Bailey D., Renouf D.V., Hounsell E.F.;
 RT "Analysis of the glycosylation patterns of the extracellular domain of
 RT the epidermal growth factor receptor expressed in Chinese hamster
 RT ovary fibroblasts."
 RL Growth Factors 13:121-132(1996).
 RN [17]
 RP CARBOHYDRATE-LINKAGE SITES ASN-56; ASN-352; ASN-361; ASN-568 AND
 RP ASN-603.
 RX MEDLINE=20198209; PubMed=10731668;
 RA Sato C., Kim J.-H., Abe Y., Saito K., Yokoyama S., Kohda D.;
 RT "Characterization of the N-oligosaccharides attached to the atypical
 RT An-X-Cys sequence of recombinant human epidermal growth factor
 RT receptor."
 RL J. Biochem. 127:65-72(2000).
 RN [18]
 RP PARTIAL SEQUENCE, AND DISULFIDE BONDS.
 RX MEDLINE=98225196; PubMed=9556602;
 RA Abe Y., Odaka M., Inagaki F., Lax I., Schlessinger J., Kohda D.;
 RT "Disulfide bond structure of human epidermal growth factor receptor."
 RL J. Biol. Chem. 273:11150-11157(1998).
 RN [19]
 RP REVIEW
 RX MEDLINE=87297456; PubMed=3039909;
 RA Carpenter G.;
 RT "Receptors for epidermal growth factor and other polypeptide
 RT mitogens."
 RL Annu. Rev. Biochem. 56:881-914(1987).
 CC -I- FUNCTION: Receptor for EGF, but also for other members of the EGF
 CC family, as TGF-alpha, amphiregulin, betacellulin, heparin-binding
 CC EGF-like growth factor, Gp30 and vaccinia virus growth factor. Is
 CC involved in the control of cell growth and differentiation.
 CC -I- FUNCTION: Isoform 2/truncated isoform may act as an antagonist.
 CC -I- CATALYTIC ACTIVITY: ATP + a protein tyrosine - ADP + protein
 CC tyrosine phosphate.
 CC -I- SUBCELLULAR LOCATION: Type I membrane protein. Isoform 2 is
 CC secreted.
 CC -I- ALTERNATIVE PRODUCTS: 4 isoforms; 1/p170 (shown here), 2/p60/
 CC truncated isoform/TEGFR, 3/p110 and 4; are produced by
 CC alternative splicing.
 CC -I- TISSUE SPECIFICITY: Expressed in placenta. Isoform 2 is also
 CC expressed in ovarian cancers.
 CC -I- MISCELLANEOUS: Binding of EGF to the receptor leads to
 CC dimerization, internalization of the EGF-receptor complex,
 CC induction of the tyrosine kinase activity, stimulation of cell DNA
 CC synthesis, and cell proliferation.
 CC -I- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X00588; CAA25240.1; -;
 DR EMBL; U95089; AAB53063.1; -;
 DR EMBL; U48722; AAC50802.1; -;
 DR EMBL; U48723; AAC50804.1; -;
 DR EMBL; U48724; AAC50796.1; -;
 DR EMBL; U48725; AAC50797.1; -;
 DR EMBL; U48726; AAC50798.1; -;
 Query Match 46.5%; Score 3166; DB 1; Length 1210;
 Best Local Similarity 49.8%; Pred. No. 4.3e-161;
 Matches 630; Conservative 177; Mismatches 352; Indels 106; Gaps 21;
 QY 11 LIALALLPPGAA--STQVCTGTDMLRLPASPEFHLMDRLHYOGCQVQGNLELTLPVN 69
 DB 14 LLAALCPASRALEBKVCQTSNKLTLQGLTFEDHFLSLQRMFNNECVLGNLEITYVQRN 73

QY 69 ASLSFLQDIQEVGYVLIHAHQVQVPLQRLRIVRGTLQFEDNYALAVLDNGDPLNNTTP 128
 DB 74 YDLSFLKTIQEVAGYVLIALNTVERIPLENLQIRGNMYVENSALAVLSNYD----- 136
 QY 129 VTGASPGGIRELQRLSLTEILKGGVLIQRNPOLCYODTILWKDIFKHNQLALTLIDTNR 188
 DB 127 ---ANKTGLKELPMRNLQELHGAVRFSNNPALCNVESIQWRDIVSDVLSNMDPQNH 183
 QY 189 SRACHPCSPWCKSGRCWGSSESDCQSLTTRTVACAGCA-RCKGPLPTDCHEQCAAGCTGP 247
 DB 184 LSCQCKDPSFCNPGSCWGAENCKLTKIIQAQCSGRCRKGKSPSCDCHNQCAAGCTGP 243
 QY 248 KHSDDLACIHNHSGICELHCPALVYNTTDTFESMPNPEGRYTFGASCVTACPYNYLSTD 307
 DB 244 RESDCLVCRKFRDEATCKDTPPLMLNPTTYQMDVNPGEKYSFGATCKVKCPRNVTVD 303
 QY 308 VGSCTLVCPLNQEVTAEDGTORCEKCKPCARVCYGLGMEHLRREVRVTSANIQFAGC 367
 DB 304 HGSCVRACGADSYEM-EEDGVKCKCKCGCRKVCNGIGIGEFKDSLSINATNIKFKNC 362
 QY 368 KKFISLAFLPESFUDGPASNTAPLOEQLOVFETLEEITGYLYISAWPDSLDPDLVFQON 427
 DB 363 TSSISGDLHILPVAFRGDSFTHTPPDQELDLKTVKEITGFLIOAWFENRTDLHAFEN 422
 QY 428 LQVIRGRILHNGAYSILTQGLGISWLSRLSRELGSGLALIHHTHLCFVHTVPMQDLFR 487
 DB 423 LEIRGRTKQHQFSLAVVSLNITSLGLRSLKLEISDGDVVISGNKNLCYANTINWKKLFG 482
 QY 488 NPHQALLHTANPEDECVEGEGLAGHOLCARGHCWGPGTQCVCNCSOFLRGQECVSCRVL 547
 DB 483 TSGQKTKIISNRGENSKATGQVCHALCSPGCGPEPRDCVSCRNVSRGECVDCNLL 542
 QY 548 QGLPREYVNAHCLPCHPCCQNGSVTCFGEADOCVACAHKDPFCVARGCPSVKPD 607
 DB 543 EGEPRFVENSECICHPECLPQAMNITCTGRGPDNCIOCAHYIDGPHCVKTCFAGVMGE 602
 QY 608 LSYMPITWKEFPDEGACQPCFINCTHSCVDLDKGCFAEQRASPLTSIVSAVVG---ILLV 664
 DB 603 NNTL-VKMYADAGHVCHLCPNCTGYCTGPGLEGCTNGPKIP--SIATGMVGAALLLV 659
 QY 665 VLVGVVGLIIRROOKIRKYMRLLQETELVEPLTPSGAMPNQAOIRLKEETURKVK 724
 DB 660 VALGIG---LFWRRRHRIVKRTLRLQLQERELVEPLTPSGEAPNQALLRLKETEFKKIK 716
 QY 725 VLGSAGFTVYKGIWIPDCENVKIPVAIKVLRNTSPKANKETLDEAYVMAGVGSVYVR 784
 DB 717 VLGSAGFTVYKGLWIPGEKVKIPVAIKELREATSPKANKETLDEAYVMASVDNPHVCR 776
 QY 785 LLGICLTSTVQLVTQIMPYGCLLDHVRENRRGLSGQDLNLCWQIAKGMYSYLEDVRLVHR 844
 DB 777 LLGICLTSTVQLITQLMPFGCLLDVREHKDNIGSOYLLNWCQIAKGMNILEDRVLVR 836
 QY 845 DLAAARNLVKSNHVKITDFGLARLLDIDETHYADGGKVPKKNMALESILRRFRTHQSD 904
 DB 837 DLAAARNLVKTPQHKITDFGLAKLGLAEEYHAEKGVPIKKNMALESILHRIYTHQSD 896
 QY 905 VMSYGVYVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTDIVYMWKCMWIDSE 964
 DB 897 VMSYGVYVWELMTFGSKPYDGIIPASEISILEKGERLPQPPICTDIVYMWKCMWIDAD 956
 QY 965 CRPRFRELVESEPRMARDPQRFWIQQ-NEDLGPASPLDSTFYRSLLEDMDGDLVDAEY 1023
 DB 957 SRPKFRELIIIEFSKWARDPQRYLVILQGDHMLPSTDSNFYALMDEEDMDVDVADY 1016
 QY 1024 LVPQCGFFCPDPAPGAGGMVHRHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEGAS 1083
 DB 1017 LIPQCGFF-----SSPSTSRTPLLSSLSATS 1042
 QY 1084 DYFDGDLGMAAKGLQSLPHTDPSPLORYSEDPVPLPSET--DGYVAPLTCSPQPEYVN 1141
 DB 1043 N--NSTVACIDRNLGQSCPIKEDSFLQRYSSDPTGALTEDSIDDTPL-----PVPEYN 1094

```
QY 1142 QDVRPOPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKDVFAGGAVENPEYL-TPQ 1200
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1095 Q-SVPRKPGASQVNVHNOPLNP-----APSRDPHYQD--PHSTAVCNPEYLNWQ 1143
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1201 GGAAPQHPPPAFSPAPDNLNYYWQ-----DP-----PERGAPSTFKGPTTAE 1244
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1144 -----PTCVNSTFDSAPHAQAQSGHQISLDNPDYQDQDFPKKAPNGIFKGS-TAE 1193
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1245 NPEYL 1249
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1194 NAEYL 1198

RESULT 5
EGFR_MOUSE STANDARD; PRT; 1210 AA.
AC Q01279;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112).
GN EGFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Liver;
RX MEDLINE=93026370; PubMed=1408137;
RA Avivi A., Skorecki K., Yayon A., Givol D.;
RT "Promoter region of the murine fibroblast growth factor receptor 2
(bek/KGFR) gene.";
RL Oncogene 7:1957-1962(1992).
[2]
RN SEQUENCE FROM N.A.
RC STRAIN=BALB/c, and CD-1; TISSUE=Liver, and Decidua;
RX MEDLINE=93126380; PubMed=7678348;
RA Paria B.C., Das S.K., Andrews G.K., Dey S.K.;
RT "Expression of the epidermal growth factor receptor gene is regulated
in mouse blastocysts during delayed implantation.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:55-59(1993).
[3]
RN SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Liver;
RA Hibbs M.L.;
RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
[4]
RN SEQUENCE FROM N.A.
RC STRAIN=B6/C3; TISSUE=Liver;
RX MEDLINE=94170986; PubMed=8125255;
RA Luetkeke N.C., Phillips H.K., Qiu T.H., Copeland N.G., Earp H.S.,
RA Jenkins N.A., Lee D.C.;
RT "The mouse waved-2 phenotype results from a point mutation in the EGF
receptor tyrosine kinase.";
RL Genes Dev. 8:399-413(1994).
[5]
RN SEQUENCE OF 1-714 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=91232866; PubMed=2030916;
RA Avivi A., Lax I., Ullrich A., Schlessinger J., Givol D., Morse B.;
RT "Comparison of EGF receptor sequences as a guide to study the ligand
binding site.";
RL Oncogene 6:673-676(1991).
[6]
RN SEQUENCE OF 969-1117 FROM N.A.
RC STRAIN=C3H;
RA Bisinger D.P., Serrero G.;
RL Submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF,
AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND
VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
(PHOSPHORYLATION (BY PKC) (BY SIMILARITY)).
PHOSPHORYLATION (AUTO-) (BY SIMILARITY)).
```

```
CC tyrosine phosphate.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- MISCELLANEOUS: Binding of EGF to the receptor leads to
CC dimerization, internalization of the EGF-receptor complex,
CC induction of the tyrosine kinase activity, stimulation of cell DNA
CC synthesis, and cell proliferation.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X78987; CAA5587.1; -
CC EMBL; U03425; AAA1789.1; -
CC EMBL; X59698; CAA42219.1; -
CC EMBL; L06864; AAA53029.1; -
CC EMBL; Z12608; CAA78249.1; -
CC HSSP; P11362; IFGK.
CC -----
CC MGD; MGI:95294; Egfr.
CC InterPro; IPR000494; EGFR_L domain.
CC InterPro; IPR000719; Euk_pkinase.
CC InterPro; IPR002174; Furin-like.
CC InterPro; IPR001245; Tyr_pkinase.
CC Pfam; PF00069; pkinase; 1.
CC Pfam; PF00757; Furin-like; 1.
CC Pfam; PF01030; Recep_L domain; 2.
CC ProDom; PD000001; Euk_pkinase; 1.
CC SMART; SM00261; FU_3.
CC SMART; SM00219; TyKc; 1.
CC PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
CC PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
CC PROSITE; PS0011; PROTEIN_KINASE_DOM; 1.
CC Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
CC Tyrosine-protein kinase; ATP-binding; Phosphorylation; Repeat.
CC SIGNAL 1 24 POTENTIAL.
CC CHAIN 25 1210 EPIDERMAL GROWTH FACTOR RECEPTOR.
CC DOMAIN 25 647 POTENTIAL.
CC TRANSMEM 648 670 EXTRACELLULAR (POTENTIAL).
CC DOMAIN 671 1210 CYTOPLASMIC (POTENTIAL).
CC REPEAT 75 300 APPROXIMATE.
CC REPEAT 390 600 APPROXIMATE.
CC DOMAIN 1028 1071 SER-RICH.
CC NP_BIND 720 728 ATP (BY SIMILARITY).
CC BINDING 747 747 ATP (BY SIMILARITY).
CC ACT_SITE 839 839 BY SIMILARITY.
CC DISULFID 190 199 BY SIMILARITY.
CC DISULFID 194 207 BY SIMILARITY.
CC DISULFID 215 223 BY SIMILARITY.
CC DISULFID 219 231 BY SIMILARITY.
CC DISULFID 232 240 BY SIMILARITY.
CC DISULFID 236 248 BY SIMILARITY.
CC DISULFID 251 260 BY SIMILARITY.
CC DISULFID 264 291 BY SIMILARITY.
CC DISULFID 295 307 BY SIMILARITY.
CC DISULFID 311 326 BY SIMILARITY.
CC DISULFID 329 333 BY SIMILARITY.
CC DISULFID 506 515 BY SIMILARITY.
CC DISULFID 510 523 BY SIMILARITY.
CC DISULFID 526 535 BY SIMILARITY.
CC DISULFID 539 555 BY SIMILARITY.
CC DISULFID 558 571 BY SIMILARITY.
CC DISULFID 562 579 BY SIMILARITY.
CC DISULFID 582 591 BY SIMILARITY.
CC DISULFID 595 617 BY SIMILARITY.
CC DISULFID 620 628 BY SIMILARITY.
CC DISULFID 624 636 BY SIMILARITY.
CC MOD_RES 680 680 BY SIMILARITY.
CC MOD_RES 1092 1092 PHOSPHORYLATION (BY PKC) (BY SIMILARITY).
PHOSPHORYLATION (AUTO-) (BY SIMILARITY)).
```

```

FT MOD_RES 1110 1110 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1172 1172 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1197 1197 PHOSPHORYLATION (AUTO-, MAJOR SITE)
      (BY SIMILARITY).
FT CARBOHYD 128 128 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 175 175 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 352 352 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 413 413 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 444 444 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 528 528 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 568 568 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 603 603 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 19 19 C -> S (IN REF. 2).
FT CONFLICT 539 539 C -> W (IN REF. 5).
FT CONFLICT 991 991 L -> F (IN REF. 4).
FT CONFLICT 1116 1117 HP -> DR (IN REF. 6).
SQ SEQUENCE 1210 AA; 134853 MW; 690E20D46DF2D2F5 CRC64;

Query Match
Best Local Similarity 49.8%; Pred. No. 5,7e-160;
Matches 633; Conservative 170; Mismatches 359; Indels 110; Gaps 23;

QY 11 LLLALLPPGAA--STQVCTGTDMLRLPASPETHLMLRLHYOGCQVQGNLELYLPTN 68
DB 14 LLLTCAAGALBEKKVCOGTSNRLTQGTGFEDHFLSLQRMNNECEVVLGNLEITYQVRN 73
QY 69 ASLSFLQDIOEVQGYLIASHNOVQVPLQRLIRVRGTOLFEDNYVALAVLNDGDPNNVTP 128
DB 74 YDLSFLKTIQEVAGVLIANTVERIPLENLQIRGNALYENYVALAILSN-----124
QY 129 VTGASPGGLRELQRLSRLTEILKGGVLIQPNPOLCYQDTILWKDI----FKHNNQLALTLI 184
DB 125 -YGNRTGLRELPMRLNQLIELIAGVRSNNPILCNMDTIQWRDIVQNVFNSMDL---180
QY 185 DTRSRACHPCSPMCKSGRCWGSSEDCQSLRTVTCAGGCA-RCKGPLPTDCHEQCAAG 243
DB 181 -QSHSPCKPCDPCPCNGSCNGSGGGEENCQRLTKIIAQQCQSHRCGRSPSDCCNOCAG 239
QY 244 CTGPKHSDDLACIENHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTAPYNY 303
DB 240 CTGPRSDCLVCKQFQDEATKDTCTPLMLYNTTYQMDVNPBGKVSFGATVKKCPRNY 299
QY 304 LSTDVGSCTLVCLPHNOEVTAEIDGTORCEKSCPCARVCYGLGMLHREVRVAVTSANIQE 363
DB 300 VVTDHGSVCRAQDPDYEV-EEDGIRKCKDCGCRKVCNGIGIGBEKDTLSINATIKH 358
QY 364 FAGCKIFGSLAFPSFDGDPASNTAPLOEQVFPETLEETGYLYISAWPDSLPDLS 423
DB 359 FKCYTAISGDLHLPVAFKGDSTRTPLDPRELEILKTVKEITGFLIIQAWPDNWDLH 418
QY 424 VFONLQVIRGRLHNGAYSILTLOGLGSWGLSRLSRLSGSLALIHNNTHLCFHVHPWD 493
DB 419 AFENLEIIRTKOHGQFSLAVVGLNLTSLGLSLKEISDGVIIISGNRLCYANTINWK 478
QY 484 QLFERNPHQALLHTANRPEDECVGEGLAGHACHGCGPGTQCVCNCSQFLRGQECVEE 543
DB 479 KLFGTNQTQKIMNNAEAKDCKAVHVCNPLCSGCGCGPEPRDCVSCQNVSGRECEVK 538
QY 544 CRVLQGLPRYVNAHCLPCHPCQPNQSVTCFGEADQCVCAHYKDPFPCVACPSG 603
DB 539 CNILEGEPREFVENSECIIQCHPECLPQAMNITCTGRGPDNCIQAHYIDGPHCVKTCAG 598
QY 604 VKPDLSPMTWPKFDEGACQPCPINTCHSCVDLDDKGCFAEORASPLTISVSAVGLL 663
DB 599 IMGENNTL-VWKYADANNVCHLCHACTYCACGPGLOGCEVWPSPGPKIPSIATGIVGGL 657
QY 664 VVVLGVVFGI-LIKRROOKIRKVTMRLLQETELVEPLTPSGAMPNQAOMRIKTELK 722
DB 658 FIVV-VALGICLFWRRRHVVKTLRLRLQRELVEPLTPSGAPNQAHLRIKTEFEKK 716
QY 723 KVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRENTSPTKANKEILDEAYVMAGVSPY 782

```

```

DB 717 IKVLGSGAGFTVYKGLWIPGEGKVKIPVAIKELREATSPKANKEILDEAYVMASVDNPHV 776
QY 783 SRLIGCLTSTVOLVTQMLPYPGCLLDHVRNRRGLSQDILLNMCWQIAGKMSYLEDVRLV 842
DB 777 CRLLGCLTSTVOLITQMLPYPGCLLDYVREKDNIGSQVLLNWCVOIAGKMYLEDRLV 836
QY 843 HRDLAARNVLVKSPPNHVKITDFGLARLLDIDETFEYHADGKGVPIKWMALLESILRRRPTHQ 902
DB 837 HRDLAARNVLVKTPOHKITDFGLAKLLGAEEKEYHAEGGKVPKWMALLESILHRIYTHQ 896
QY 903 SDVMSYGVTVWELMTFGAKPYDIPAREIPDLLEKGERLPPORPCTIDYVMIMVKCMID 962
DB 897 SDVMSYGVTVWELMTFGSKPYDIPASDISILLEKGERLPPORPCTIDYVMIMVKCMID 956
QY 963 SECRRPRELVSFSSMARDPQRFVVIQ-NEDLGPASPLDSTFYRSLLEDDEDDGLVDAAE 1021
DB 957 ADSRPFERELILEFSSMARDPQRYLVIQDERMHLPSPTDSNFYRALMBEEDMEDVVDAD 1016
QY 1022 EYLVPOQGFPCDPAPGAGGMVHRRSSSTRSGGDLTLGLEPSEEAAPRSPAPSEGA 1081
DB 1017 EYLVPOQGF-----NSPST-----SRTLLSSLSA 1042
QY 1082 GSDVFDGDLGMAAKGLOSLPHTDPSPLQRYSEDPTVPLPSET--DCYVAPLTCPOPEY 1139
DB 1043 TSN-----NSTVACINRNGSCRKEDAFLOQRYSDPTGAVTEONIDDAFL-----VPVEY 1092
QY 1140 VNOPDVRPOPSPREGPLPAAPAGATLERAKTILSPKNGVVKDVAFGAGAVENPEYL-T 1198
DB 1093 VNO-Q-SVPKPPAGSVQNPVHNQPLHP-----AFGRDLHYQN--PHSNVAGNPEYLVNT 1141
QY 1199 PQGGAAPPHPPAFSPAFDNLYWDQ-----DP-----PERGAPPSTFKGTPT 1242
DB 1142 AQ-----PTCLSSGSGSPALWIOKGSHQWMLNDPNYQDDFFPKPKNGIFKNG-PT 1191
QY 1243 AENPEYLGIDVP 1254
DB 1192 AENAEYLRVAPP 1203

RESULT 6
ERB4_HUMAN STANDARD; PRT; 1308 AA.
ID ERB4_HUMAN
AC Q15303;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112)
DE (p180erbB4) (tyrosine kinase-type cell surface receptor HER4).
GN ERB4 OR HER4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A. (ISOFORM JM-A).
RP TISSUE=Breast carcinoma;
RX MEDLINE=93189574; PubMed=8383326;
RA Plowman G.D., Culouscou J.-M., Whitney G.S., Green J.M., Carlton G.W.,
RA Foy L., Neubauer M.G., Shoyab M.;
RT "Ligand-specific activation of HER4/p180erbB4, a fourth member of the
RT epidermal growth factor receptor family.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:1746-1750(1993).
RN [2]
SEQUENCE FROM N.A. (ISOFORMS JM-A AND JM-B).
RP TISSUE=Fetal brain;
RX MEDLINE=97476287; PubMed=9334263;
RA Elenius K., Corfas G., Paul S., Choi C.J., Rio C., Plowman G.D.,
RA Klagsbrun M.;
RT "A novel juxtamembrane domain isoform of HER4/erbB4. Isoform-specific
RT tissue distribution and differential processing in response to
RT phorbol ester.";
RL J. Biol. Chem. 272:26761-26768(1997).

```

FT	DISULFID	308	323	BY SIMILARITY.
FT	DISULFID	326	330	BY SIMILARITY.
FT	DISULFID	503	512	BY SIMILARITY.
FT	DISULFID	507	520	BY SIMILARITY.
FT	DISULFID	523	532	BY SIMILARITY.
FT	DISULFID	536	552	BY SIMILARITY.
FT	DISULFID	555	569	BY SIMILARITY.
FT	DISULFID	559	577	BY SIMILARITY.
FT	DISULFID	580	589	BY SIMILARITY.
FT	DISULFID	593	614	BY SIMILARITY.
FT	DISULFID	617	625	BY SIMILARITY.
FT	DISULFID	621	633	BY SIMILARITY.
FT	MOD_RES	1162	1162	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	MOD_RES	1188	1188	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	MOD_RES	1258	1258	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	MOD_RES	1284	1284	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	CARBHYD	138	138	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBHYD	174	174	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBHYD	181	181	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBHYD	253	253	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBHYD	358	358	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBHYD	410	410	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBHYD	473	473	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBHYD	495	495	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBHYD	548	548	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBHYD	576	576	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBHYD	620	620	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	VARSPLIC	626	648	NGPTSHDCIYVPTGHSTLPQHA -> IGSIEDICIGLMD (IN ISOFORM JM-B).
FT	SQ	SEQUENCE	1308 AA; 146807 MW; 584AE80985D88761 CRC64;	
	Query Match	44.1%;	Score 3003.5; DB 1; Length 1308;	
	Best Local Similarity	45.6%;	Pred. No. 2.le-152;	
	Matches	614; Conservative	184; Mismatches 375; Indels 173; Gaps	28;
Qy	9 WGLLLALLPGAA-----STOVCTGTMKRLRPASPETHLDMLRHLRYOGCVOVGNNLELT	64		
Db	8 WVVSLVLVAAGTVQPDSQSVCAGTENKLSSLDLEQQYRALRKYYENCEVWGNLEITS	67		
Qy	65 LPTNASLSFLDIOEQGYVLIHAHQVRVPQLRLRVGTQTLPEDNYALAVLDNGPLN	124		
Db	68 IEHRDLISFLRSREVTGYVLVALNGFRYLPLENRRIIRGKTLYEDRYALAFILNYSKDG	127		
Qy	125 NTPPVGTASPGGLRELQLRLSRLTBILKGGVLIORNPOLCYODTILWKDI PHKNQALTLI	184		
Db	128 NF-----GLQEGLKMLTEILNGGVVDQNKFCLYADTTTHWQDIVRNPFNSUTLV	178		
Qy	185 DTNRSRACHPCSPMKGRSGWBESSDCQSLRTITVCAGGC-ARKCGPLPTDCHEQCAAG	243		
Db	179 STNGSSGCRCHKSCTG-RCMGPTENHCQTLLTRTVCAEQCDGRCYGPVSDCCCHRECAGG	237		
Qy	244 CTGPKHSIDLACLHFNHSGICEUHCAPALVTYNTDTTFESMPNPGRYTFGASCVTACPYNY	303		
Db	238 CSGFKDTCDFACMNFNDSCAFCVQCPTFYFNPTTFFOLEHNFAKTYGAFCKVKCPHF	297		
Qy	304 LSTDVGSCTLVCPLNHOEVTAEDGTORCEKSKPCARCVCYGLGWEHLREVRAVTSANIQE	363		
Db	298 V-VDSSSCVTRACSSXMEV-EENGIMCKEPCETDICPAACGDIGTGSLMSAQTVDDSSNIDK	355		
Qy	364 FAGCKIFGSLAPFPESFDGDPASNTAPLOPEQLQVFTEEITGLYIISAWPDSLDPDS	423		
Db	356 FINCTKINGNLIELVTGIHGDPYNAIEADPEKLVNVRTVREITGTFLNIQSWPPNMWDFS	415		
Qy	424 VFONLQVIRGRILNHGAYSILTUGLGISWLGRSLRELGSGLALIIHHNTHLCFHVITVPWD	483		
Db	416 VFSNLVTIGRVLVYGLSLILLKQQGITSIQFOSLRKEISAGNTYITDNSNLCYHTTNWT	475		
Qy	484 QLFERNPHOALLHTANRPEDECVEGLEACHOLCARGHCWGPGPTQCVNCISOFLRGQECVEE	543		
Db	476 TLFSTINQRIVIRDNRKAENCTAEGVMCNHLSSDGCWGPGPDQCSRFRRSGRICIES	535		
Qy	544 CRVLQGLPIREYVNARHCLCPHCBPCQP-QNGSVVTCFGPEADQCVCACAHYKDPPFCVARCPS	602		

```
Db      536  CNLYDGEFRFENGSI CVCDCPOCEKXEDGLLTCHGPGDNCCTKSHFKDGNPCVCKPD 595
Qy      603  GVKPDLSYMPDWKPFEBEGACQCPINCTHSCVDLDDKGC-----PABQASPL 651
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      596  GLOQANSF--IFKYADPDRECHPCPNCTCGNPTSHDCIYYPWTGSHSTLPQHAR-TPL 652
Qy      652  TSTVSAVV-CILLVVLGVVFGILIKRQOKIRKYTMRRLLQSTELVPLTPSGAMPNOA 710
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      653  --TAAGVIGGLFLLIVVGLTFVAVYRRKSIK-KRALLRRL-ETELVPLTPSGTAPNOA 708
Qy      711  QMRILKETELRKVKVLSGAGFYVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDE 770
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      709  QLRIKETELRKVKVLSGAGFYVYKGIWIPDGETVKIPVAIKILNETTGPKANVFEMDE 768
Qy      771  AYVAGVGSYVSRLLGICLTSTVLQVLTQMPYGCCLLDHVRNRLGSGDLLNWCQVIA 830
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      769  ALIMASMDHPLVRLVLLGCVLSPTIQLVTLMPHGLLEYVHEHKDIGNISQLLNWCQVIA 828
Qy      831  KGMNLYEDVLRVHRDLAARNVLKSNHVKITDPLGLARLLDIDETEHADGGKVPKQWMA 890
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      829  KGMNLYEERLVRVHRDLAARNVLKSNHVKITDPLGLARLLGDEKEYNADGGKMPKQWMA 888
Qy      891  LESILARRRTHQSDVMSYGVVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTID 950
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      889  LECIHVRKFTHQSDVMSYGVVWELMTFGKPYDGIIPAREIPDLLEKGERLPPOPICTID 948
Qy      951  VYIMVKCMWIDSECRPRFRELVSFMRMARDQRFVVIQNEP-LGPASPLDSTFYRSL 1009
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      949  VYIMVKCMWIDADSRPKFELAAEFMRMARDQRFVVIQNEP-LGPASPLDSTFYRSL 1008
Qy      1010  EDDMDGLVDAAEYLVPQGFPCDPAPGAGVHHRSSSTRSGGDLTLGLPSEEE 1069
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1009  DEEDMDMAAEYLVP-QAFNIPP-----IYTSRARIDSRS-----EIGHSPPPAY 1056
Qy      1070  APS-----PLAP--SEGAGSDVFDGLGMAAGLQ 1100
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1057  TPMGNGQVYRDCGFAAEQGVYPRAPTITIEAPVAQATAEIFDDSCNCTLKRKPA 1116
Qy      1101  LPTHDPSPQRYSEDTVPPLPS-----ETDGVAPLTCSPQPEYVNPQDVRPQPPSPR 1153
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1117  PHVQEDSSTQRYSDPTVPAPERSPRGELDEEGVMTMRKPKQEYLNRYE----- 1167
Qy      1154  EGPLPARPAGATLERAKTLPSCNGVGVKDVFAFGGAVENPEYLTPOGGAAPQHPPPA- 1212
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1168  ENPFVSR-----KNGDLQ-----ALDNPYHNASNG-----PPKAE 1199
Qy      1213  -----FSPAFONLYYWDQDPPPERGA--PPSTF 1237
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1200  DEVNBEPLYNTFANTLGKAEYLNKILSMPEKAKAFDNPDYWNHSLPRSTLQHPDYL 1259
Qy      1238  KGTFP-----AENPEYL 1249
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1260  QEYSTKYFYKQNGRIPIVAENPEYL 1285
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
RESULT 7
ID_ ERB4_RAT STANDARD; PRT; 1308 AA.
AC O62956; Q922N7;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112).
GN ERBB4 OR TYRO-2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Heart;
RX MEDLINE=98221155; PubMed=9553078;
RA Zhao Y.-Y., Sawyer D.R., Baliga R.R., Opel D.J., Han X.,
```

```
RA Marchionni M.A., Kelly R.A.;
RT "Neuregulins promote survival and growth of cardiac myocytes.
RT Persistence of ErbB2 and ErbB4 expression in neonatal and adult
RT ventricular myocytes.";
RL J. Biol. Chem. 273:10261-10269(1998).
RN [2]
RP SEQUENCE OF 848-901 FROM N.A.
RC TISSUE=Sciatic nerve;
RX MEDLINE=91222560; PubMed=2025425;
RA Lai C., Lemke G.;
RT "An extended family of protein-tyrosine kinase genes differentially
RT expressed in the vertebrate nervous system.";
RL Neuron 6:691-704(1991).
RN [3]
RP SEQUENCE OF 1031-1198 FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Spinal cord;
RX MEDLINE=97184212; PubMed=9030624;
RA Carroll S.L., Miller M.L., Frohnert P.W., Kim S.S., Corbett J.A.;
RT "Expression of neuregulins and their putative receptors, ErbB2 and
RT ErbB3, is induced during Wallerian degeneration.";
RL J. Neurosci. 17:1642-1659(1997).
CC -!- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULINS, NRG-
CC 2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND
CC NTAK. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.
CC NOT ACTIVATED BY EGF, TGF-A, AND AMPHIREGULIN (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
CC RECEPTORS (POTENTIAL).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN THE DEVELOPING
CC NERVOUS SYSTEM. EXHIBITS DISTINCT AND HIGHLY REGIONALIZED PATTERNS
CC OF EXPRESSION IN THE ADULT BRAIN, WHERE IT IS MAINLY FOUND IN THE
CC RETICULAR NUCLEUS OF THE THALAMUS. VERY LOW LEVELS IN KIDNEY, AND
CC HEART.
CC -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF041838; AAD08899.1; -.
CC EMBL; U52531; AAC53051.1; -.
CC HSSP; P11362; 1FGK.
CC InterPro; IPR000494; EGFR_L domain.
CC InterPro; IPR000719; Euk_pkinase.
CC InterPro; IPR002174; Furin-like.
CC InterPro; IPR001245; Tyr_pkinase.
CC InterPro; IPR004019; YLP motif.
CC Pfam; PF00757; Furin-like; 1.
CC Pfam; PF00069; pkinase; 1.
CC Pfam; PF01030; Recep_L domain; 2.
CC Pfam; PF02757; YLP_2.
CC PRINTS; PR00109; TYRKINASE.
CC ProDom; PD000001; Euk_pkinase; 1.
CC ProDom; PD000261; FU; 4.
CC SMART; SM00219; TyrKc; 1.
CC PROSITE; PS00107; PROTEIN KINASE ATP; 1.
CC PROSITE; PS00111; PROTEIN KINASE_TYR; 1.
CC PROSITE; PS00109; PROTEIN_KINASE_DOM; 1.
CC Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
CC Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 1308 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-4.
FT DOMAIN 26 651 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 652 675 POTENTIAL.
FT DOMAIN 676 1308 CYTOPLASMIC (POTENTIAL).
```

```
FT DOMAIN 186 334 CYS-RICH.
FT DOMAIN 496 633 CYS-RICH.
FT DOMAIN 718 985 PROTEIN KINASE.
FT NP BIND 724 732 ATP (BY SIMILARITY).
FT BINDING 751 751 ATP (BY SIMILARITY).
FT ACT SITE 843 843 BY SIMILARITY.
FT DISULFID 189 197 BY SIMILARITY.
FT DISULFID 193 205 BY SIMILARITY.
FT DISULFID 213 221 BY SIMILARITY.
FT DISULFID 217 229 BY SIMILARITY.
FT DISULFID 230 238 BY SIMILARITY.
FT DISULFID 234 246 BY SIMILARITY.
FT DISULFID 249 258 BY SIMILARITY.
FT DISULFID 262 289 BY SIMILARITY.
FT DISULFID 293 304 BY SIMILARITY.
FT DISULFID 308 323 BY SIMILARITY.
FT DISULFID 326 330 BY SIMILARITY.
FT DISULFID 503 512 BY SIMILARITY.
FT DISULFID 507 520 BY SIMILARITY.
FT DISULFID 523 532 BY SIMILARITY.
FT DISULFID 536 552 BY SIMILARITY.
FT DISULFID 555 569 BY SIMILARITY.
FT DISULFID 559 577 BY SIMILARITY.
FT DISULFID 580 589 BY SIMILARITY.
FT DISULFID 593 614 BY SIMILARITY.
FT DISULFID 617 625 BY SIMILARITY.
FT DISULFID 621 633 BY SIMILARITY.
FT MOD RES 1162 1162 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD RES 1188 1188 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD RES 1258 1258 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD RES 1284 1284 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 138 138 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 174 174 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 253 253 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 358 358 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 410 410 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 473 473 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 495 495 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 548 548 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 620 620 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 1062 1062 S -> N (IN REF. 3).
FT CONFLICT 1080 1082 PYT -> SYR (IN REF. 3).
SQ SEQUENCE 1308 AA; 146957 MW; D944BB0996A08B41 CRC64;

Query Match
Best Local Similarity 43.88; Score 2984; DB 1; Length 1308;
Matches 612; Conservative 191; Mismatches 384; Indels 160; Gaps 28;

Qy 1 MELA-ALCRWGLLL--ALLPFGAASCTGCTGDMKRLPASPTHLDMLRHLRYGCGVQV 57
Db 1 MKLATGLVWVGSLLVAARTVQPSASQSCVAGTENKLSLSLDLEQQVRLARKYVENCVM 60

Qy 58 GNLELTVLPNASLSFLDIOEVQGVYLIHNRQVQVLPQRLRIVRGTLQFEDNYALAVL 117
Db 61 GNLEITSIEHNRDLISFLRSIREVTGYVVALNQFRVLPLENLRIIRGTLKYEDRYALAI 120

Qy 118 DNGDPLNNTTPTVTCASPGGLRELQLRSLTEILKGGVLIQRNPOLCYQDTILKWDIFHKN 177
Db 121 LNYRKQGNF-----GLQELGKLNLTIELNGVYVQNKFLCYADTIHWQDIVRNPW 171

Qy 178 QLALTLDITNRSRACHPCSPMKGSRCSGSSDQSLTRTCVAGGC-ARCKGPLPTDCC 236
Db 172 PSNMTLVSTIGSSGCGRCHSKCTG-RCWGPTEHNCQTILFTVCABQDCGRCYGYVSDCC 230

Qy 237 HEQCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCV 296
Db 231 HRECAAGCGSKPOTDCFACMNFNDSGACVTCQCPQTFVYNPTTTFQLEHNFNAKYTYGAFV 290

Qy 297 TACPYNVLSLDVGSCTLVCPHINQEVTAEDGTORCEKSPCARVCYGLGMEHLREVRVAV 356
Db 291 KKCPHNFV-VDDSSCRVACPSSKMEV--EENGIRCKPCTDIDCPKACDGGTGTGSLMSAQTV 348
```

RESULT 8
XMRK_XIPMA
ID XMRK_XIPMA
AC P13388;

STANDARD; PRT; 1167 AA.

```
Qy 357 TSANIOEFAGCKKIFGSLAFIPESFDGDPASNTAPLQPEQLQVFETLEBITGYLISAWP 416
Db 349 DSSNIDKFNCTKINGNLIFLVGTIIGDPYNAIDAIDPEKJNVFRIVREITGLNTQTPW 408

Qy 417 DSLPDLVSFQNLQVIRGRILHNGAYSLLTQGLIGISWLGRLRSRELGSGLALIHNNHLCF 476
Db 409 PNMTDFSVSNLVTIGGRVLYSGLSLLILKQQGITSIQFQSLKEISAGNIYITDNSLCY 468

Qy 477 VHTVPWDLFRNPHOALLHTANRPEDECVGEGLACHQLCARGHCWGGPTQCVNCVQFLR 536
Db 469 YHTINWTTLFTSTVNQRIVIRDNRAENCTAEGMVNHLCSNDGCGPGDPQCLSCRRFSR 528

Qy 537 GQECVEECRVLOGLPREYVNAHCLPCHPEOP-QNGSVTCFPEADQCVACAHYKDPFF 595
Db 529 GKICIESCNLYDGEFEPENGSIQVECDQCEKMEGLTCHGPGPDNCTKCSHFQDGN 588

Qy 596 CVARCPGKVPDLSYMPIWKFPDEBAGCQPCINCTHSCVDDLDKGC-----PA 644
Db 589 CVEKCPDVLQGANSF--IFKYADQDRECHPCPNCTQGCNGPTSHDCIYYPWTGHSTLPQ 646

Qy 645 EQRASPLTSIVSAVV-GILLVVVLGVVGLILKRRQOKIRKYTMRELLLOETELVEPLTPS 703
Db 647 HAR-TPL--IAAGVIGGLFILVIMALTFAVYVRRSIK-KKRALRFL-ETELVEPLTPS 701

Qy 704 GAMPNQAQMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVRENTSPKA 763
Db 702 GTAPNQAQRLILKETELKEVKVLGSGAFGTVYKGIWVPEGETVKIPVAIKILNETGPKA 761

Qy 764 NKEILDEAYMAGVGSPPYVSRLLGICLTSTVOLVQLMPYGCILLDHVRNRLGSGQDLL 823
Db 762 NVFEMDEALIMASVDHPHLVRLLVGLVCLSPITQVLTQMPHGCLLEYVHEHKONIGSOLL 821

Qy 824 NNCMOIQAKMSVLEDVLRVLRDLAARNVLKSPNHVKITDFGLARLLDDETEYHADGGK 893
Db 822 NNCVQIAKGMVLEERRVLRDLAARNVLKSPNHVKITDFGLARLEGEDEKEYNADGGK 891

Qy 884 VPKWMALESILRRRTHQSDVMSYGVTVVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQ 943
Db 882 MPKWMALBCHYRKFTHQSDVMSYGVTVVWELMTFGKPYDGIPTREIPDLLEKGERLPQ 941

Qy 944 PICTIDVIMVWKMWIDSECRPRFRELVSFBSRMARDPQRFVVIQNEB-LGPASPLDS 1002
Db 942 PICTIDVIMVWKMWIDADSRRPKFKELAAEFBSRMARDPQRYLVITQGDRLMKLPSPNDS 1001

Qy 1003 TFRSRLLEDDMGDLVDAEYLVPOQGFPCDP-----APGA 1039
Db 1002 KFFQNLLEDEDEDMWDAEYLVLP-QAFNIPPIYTSRTRIDSNRSEIGHSPPPAYTPMS 1060

Qy 1040 GGMVHRRHRSSTRSGGDLTLGLEPSEEEAPRSPLAPSEGAGSDVDFGDLGWGAAGLQ 1099
Db 1061 GSQFVYQDGFATQGG--MPMPYTATTSTIPEAPVA--QGATAEMFDDSCCNGTLRKPV 1115

Qy 1100 SLPTHDPSPLQYSEBPTVPLPS-----ETDGVYVAPLTCSPQPEYVNPQDVRRPQSP 1152
Db 1116 VPHVQEDSSQYRSADPTVFAPERNPRAELDEGYMTMHDKPKQBYLNPVE----- 1167

Qy 1153 REGPLPAARPACATLERAKTLSPGKNVGVKVFAGGAVENPEYLTPOGGAAPQHPHPA 1212
Db 1168 -ENPFVSR--XNGDLQ-----ALDNPEYHSASSG-----PPKA 1198

Qy 1213 -----FSPAFDNLVYMDQDPPERGA--PPST 1236
Db 1199 EDEYVNEPLYLNTFTNALGNAEYMKNSLLSVEKAKAFDNDPDYWNHSLPPRSTLQHPDY 1258

Qy 1237 FKGTPT-----AENPEYL 1249
Db 1259 LQEYSTKYFKQNGRIRPIVAENPEYL 1285
```


DT 01-JAN-1990 (Rel. 13, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Melanoma receptor protein-tyrosine kinase precursor (EC 2.7.1.112).
 GN XMRK OR TU.
 OS Xiphophorus maculatus (Southern platyfish).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
 OC Cyprinodontiformes; Poeciliidae; Xiphophorus.
 OX NCBI_TaxID=8083;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90015140; PubMed=2797166;
 RA Wittbrodt J., Adam D., Malitschek B., Maueuer W., Raulf F.,
 RA Telling A., Robertson S.M., Scharf M.;
 RT "Novel putative receptor tyrosine kinase encoded by the melanoma-
 RL inducing Tu locus in Xiphophorus.";
 RL Nature 341:415-421 (1989).
 RN [2]
 RP REVISION TO 515.
 RA Scharf M.;
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: PROBABLE RECEPTOR WITH PROTEIN-TYROSINE KINASE ACTIVITY.
 CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -!- DISEASE: INVOLVED IN PIGMENT CELLS MALIGNANT MELANOMAS.
 CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X16891; CAA34770.2; -.
 DR PIR; S06142; S06142.
 DR HSP; P11362; IFGK.
 DR InterPro; IPR000494; EGFR_L domain.
 DR InterPro; IPR000719; Euk_pkinase.
 DR InterPro; IPR002174; Furin-like.
 DR InterPro; IPR002290; Ser thr_pkinase.
 DR InterPro; IPR001245; Tyr_pkinase.
 DR Pfam; PF00069; pkinase_1.
 DR Pfam; PF00757; Furin-like; 1.
 DR Pfam; PF01030; Recep_L domain; 2.
 DR PRINTS; PD00109; TYRKINASE.
 DR ProDom; PD000001; Euk_pkinase; 1.
 DR SMART; SM00261; FU; 5.
 DR SMART; SM00220; S_TKC; 1.
 DR SMART; SM00219; TYRK; 1.
 DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
 DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.
 DR PROSITE; PS00011; PROTEIN KINASE DOM; 1.
 KW Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
 KW Tyrosine-protein kinase; ATP-binding; Phosphorylation; Proto-oncogene.
 FT SIGNAL 1 25
 FT CHAIN 26 1167 MELANOMA RECEPTOR PROTEIN-TYROSINE
 FT KINASE.
 FT DOMAIN 26 642
 FT TRANSMEM 643 665
 FT DOMAIN 666 1167 POTENTIAL.
 FT DOMAIN 710 977 CYTOPLASMIC (POTENTIAL).
 FT NP_BIND 716 724 PROTEIN KINASE.
 FT BINDING 743 743 ATP (BY SIMILARITY).
 FT ACT_SITE 835 835 BY SIMILARITY.
 FT DISULFID 195 204 BY SIMILARITY.
 FT DISULFID 199 212 BY SIMILARITY.
 FT DISULFID 220 228 BY SIMILARITY.
 FT DISULFID 224 236 BY SIMILARITY.

FT	DISULFID	237	245	BY SIMILARITY.
FT	DISULFID	241	253	BY SIMILARITY.
FT	DISULFID	256	265	BY SIMILARITY.
FT	DISULFID	269	296	BY SIMILARITY.
FT	DISULFID	300	311	BY SIMILARITY.
FT	DISULFID	315	330	BY SIMILARITY.
FT	DISULFID	333	337	BY SIMILARITY.
FT	DISULFID	504	513	BY SIMILARITY.
FT	DISULFID	508	521	BY SIMILARITY.
FT	DISULFID	524	533	BY SIMILARITY.
FT	DISULFID	537	553	BY SIMILARITY.
FT	DISULFID	556	569	BY SIMILARITY.
FT	DISULFID	560	577	BY SIMILARITY.
FT	DISULFID	593	615	BY SIMILARITY.
FT	DISULFID	618	626	BY SIMILARITY.
FT	DISULFID	622	634	BY SIMILARITY.
FT	CARBOHYD	114	114	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	144	144	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	201	201	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	356	356	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	365	365	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	398	398	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	417	417	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	501	501	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	576	576	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	621	621	N-LINKED (GLCNAC. .) (POTENTIAL).
SQ	SEQUENCE	1167 AA;	129934 MW;	4793E0749DC1D55A CRC64;

Query Match 39.9%; Score 2717.5; DB 1; Length 1167;
 Best Local Similarity 45.5%; Pred. No. 2.9e-137;
 Matches 577; Conservative 163; Mismatches 390; Indels 137; Gaps 27;

QY	4	AALCRGGLLLALLPGGAAT---	QVCTGTDMLRLPASPETHDLMLRLHYQCGVVQGN	59
DB	8	ALLQ--LLLVLISIRCCSTDPDKVCQTSNQWTM---	LDNHYLKMKMYSCNVLEN	62
QY	60	LELYPTNASLFLQDIQEVQGYLIAHNVQVPLQRLIRIVRGQTQFEDNYVALVDN	119	
DB	63	LEITYTQENQDLFLQSIQEVQGYVLIAMNEVSTIPLVNLRLTRGQNLVEGFTLVMSN	122	
QY	120	GPLNNTTPTVTGASPGRLRLQLRSLTEILKGVLIQRLNPOLCYQDTILWKDIFHKNNQ	179	
DB	123	YQK-NESSP--DVYQVGLKQLQLSNLTSLSGVSKVSHNPLLCNVETINWMDIVDKTSNP	179	
QY	180	ALTIDITNRSRACHPCSPMCKSGSRGSESDCQSLTRTVCAAGC--ARCKGPLPTCCHE	238	
DB	180	TWNLIPAFERQCKDHCVCVNGSWAPGPGHCQKFTKLLCAEQCNRCRGPRPDCNE	239	
QY	239	QCAAGCTGPKHSDCLACLHFNHSGICELHCPALVYVNTDTFESMPNPEGRYTFGASCVTA	298	
DB	240	HCAGGCTGPRATDCLACRDFNDGTCCKTCTPPPKIYDIVSHQVVDNPNIKYTFGAACVKE	299	
QY	299	CPYNLYSTDVGSCTLVCPHNGEVTAEADGTQRCCKSKPCARVCYGLGMEHLREVRVTS	358	
DB	300	CPSNYVYTE--GACVRSKCSAGMLEVD--ENGRSKCPDGVCPKVCDDIGIGLSLNTIAVNS	357	
QY	359	ANIOEPAGCKKIFGSLAFLPESFDGDPASNTAPLOPEQLQVPELLEEITGYLVISAWPDS	418	
DB	358	THIRSFNCTKINGDIIILNRNFEFGDPHYKIGTMDPEHLWNLTTVKEITGYLVIMWPNEN	417	
QY	419	LPDLSVFQNLQVIRGRILHNGAYS--LTLQGLGSIWGLRSLRGLSGLALIHHTHLCFV	477	
DB	418	MTSLSVFQNLLEIRGRTTTSRGFSFVVQVVRHLQWLGLSLRSEVSGNVLKNTLQLRYA	477	
QY	478	HTVPDQLFRNPHQALLHTANRPEDECVGEGGLACHOLCARGHCWGPGTQCVCNCSFLRG	537	
DB	478	NTINWRRLFRSEDSQSEYDART-----ENQTCNNECEDGCGWPGTMCVSCSLHVRG	530	
QY	538	QECVEECRVLOGLPREYVNVNARHCLPCHPESCPQNGSVTCFGEADOCVACAHVKDPFCV	597	
DB	531	GRCVASCNLQGEPREAQVDGRCVQCHQECQLVQTSLSLTCYGFPGPANCSSAHFQDGPQCI	590	
QY	598	ARCPGVKPDLSYMPIWKFPDDEGACQPCFINCTHSCVDLDDKGCQPAEQASPLTIVSA	657	

Db 591 PRCPHGLGDDTL-INKYADKMGQCPCHQNCQTQCGPGLSCRGD-IVSHSLAVGL 648
Qy 658 VVGILLVVLGVFGIILKRRQKIRKYVRRLLQETELVEPLTPSGAMNQOMRLIKE 717
Db 649 VSGLLIIVALLIVLLVLRRIK-RKRTIRCLQEKELVEPLTPSGQAFNAPFLRIKE 707
Qy 718 TELRKVVLGSGAGFTYKGIWIPDGENVKIPVAKVLRNTPKANKELDEAYVMAGV 777
Db 708 TEFKORVLGSGAGFTYKGLWPDGENIRIPVAKVLRNTPKANKVQEVLDAYVMASV 767
Qy 778 GSPVSVLLGICLTSTVOLVQLMPYGCCLLDHVRNENRGLSGQDLNWCQIAGKMSYLE 837
Db 768 DHPHVCRLGICLTSAVOLVQLMPYGCCLLDYVRQHQERCGQWLLNWCQIAGKMSYLE 827
Qy 838 DVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKGVPIKMALESILRR 897
Db 828 ERHLVHRDLAARNVLLKNPNHVKITDFGLSKLLTADEKEYQADGKGVPIKMALESILQW 887
Qy 898 RFTQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITDVIYIMVK 957
Db 888 TYHQSDVWSYGVTVWELMTFGSKPYDGIIPAKEIASVLENGERLPQPPICITIEVYMIILK 947
Qy 958 CWMIDSECRPRRELVSSEFMRDQRFVVIQNEIDLGASPLDSTFYRSLLEDGMDGL 1017
Db 948 CWMIDSECRPRRELVSSEFMRDQRFVVIQNEIDLGASPLDSTFYRSLLEDGMDGL -DV 1002
Qy 1018 VDAEYLVPOQGFPCPDAPGAGGVHHRSSSTRSGGDLTLGLEPSEBAPRPLAP 1077
Db 1003 VDAEYLVLPYKRI-----NRQGS-----BFCIP 1025
Qy 1078 SEGAGSDVFDGLGMAAGKGLQLPHTDPSPLOKYSDDPTV-PLPSTDCGVAPLTCSPQ 1136
Db 1026 PTH-----PVRENSITURNISDPTQNALEKDLQGH----- 1056
Qy 1137 PEYVNPQDVRPQP-----PSPRE-----GPLP-AARPAGATLERAKTLPKNGVVKD 1183
Db 1057 -EYVNPQGSSTSLSDIYNPNVEDTLGNGVPSLSQEAETNFSREYLTNTQNSL--- 1112
Qy 1184 VFAFGGAVENPEYLTPOGGAAPQHPAPAFDNLYYWDQPPERGAPPSTFKGTPTA 1243
Db 1113 PLVSSGSMDDPDY---QAG-----YQAAP-----LPQTGALTNGMFLPAA 1150
Qy 1244 ENPEYLG 1250
Db 1151 ENLEYLG 1157
RESULT 9
ID_ERB3_HUMAN STANDARD; PRT; 1342 AA.
AC P21860;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)
DE (c-erbB3) (tyrosine kinase-type cell surface receptor HER3).
GN ERB3 OR HER3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90083234; PubMed=2607875;
RA Kraus M.H., Issing W., Miki T., Popescu N.C., Aaronson S.A.;
RT "Isolation and characterization of ERB3, a third member of the
RT ERBB/epidermal growth factor receptor family: evidence for
RT overexpression in a subset of human mammary tumors.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:9193-9197(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90311312; PubMed=2164210;

RA Plowman G.D., Whitney G.S., Neubauer M.G., Green J.M., McDonald V.L.,
RA Todaro G.J., Shoyab M.;
RT "Molecular cloning and expression of an additional epidermal growth
RT factor receptor-related gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:4905-4909(1990).
RN [3]
RP SEQUENCE FROM N.A. (SHORT FORM).
RC TISSUE=Placenta;
RX MEDLINE=93282822; PubMed=7685162;
RA Katoh M., Yazaki Y., Sugimura T., Terada M.;
RT "c-erbB3 gene encodes secreted as well as transmembrane receptor
RT tyrosine kinase.";
RL Biochem. Biophys. Res. Commun. 192:1189-1197(1993).
CC -!- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTAK.
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (LONG FORM) AND
CC SECRETED (SHORT FORM).
CC -!- ALTERNATIVE PRODUCTS: TWO FORMS, A LONG TRANSMEMBRANE FORM AND A
CC SHORT SECRETED FORM, EXIST DUE TO ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: EPITHELIAL TISSUES AND BRAIN.
CC -!- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
CC -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
CC AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
CC PHOSPHATIDYLINOSITOL 3-KINASE (BY SIMILARITY).
CC -!- DISEASE: OVEREXPRESSED IN A SUBSET OF HUMAN MAMMARY TUMORS.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; M29366; AAA35790.1; -;
DR EMBL; M34309; AAA35979.1; -;
DR EMBL; S61953; AAB26935.1; -;
DR PIR; A36223; A36223.
DR HSP; P11362; IFGK.
DR Genew; HGNC:3431; ERBB3.
DR MIM; 190151; -;
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_kinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00069; kinase; 1.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR ProDom; PD000001; Euk_kinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; FALSE_NEG.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW Alternative splicing.
FT SIGNAL 1 19
FT CHAIN 20 1342
FT DOMAIN 20 643
FT TRANSMEM 644 664
FT DOMAIN 665 1342
FT DOMAIN 709 966
FT NP_BIND 715 723
FT BINDING 742 742
FT ACT_SITE 834 834
FT DISULFID 186 194
FT BY SIMILARITY.
FT 190 202
BY SIMILARITY.

FT DISULFID 210 218 BY SIMILARITY.
 FT DISULFID 214 226 BY SIMILARITY.
 FT DISULFID 227 235 BY SIMILARITY.
 FT DISULFID 231 243 BY SIMILARITY.
 FT DISULFID 246 255 BY SIMILARITY.
 FT DISULFID 259 286 BY SIMILARITY.
 FT DISULFID 290 301 BY SIMILARITY.
 FT DISULFID 305 320 BY SIMILARITY.
 FT DISULFID 323 327 BY SIMILARITY.
 FT DISULFID 500 509 BY SIMILARITY.
 FT DISULFID 504 517 BY SIMILARITY.
 FT DISULFID 520 529 BY SIMILARITY.
 FT DISULFID 533 549 BY SIMILARITY.
 FT DISULFID 552 565 BY SIMILARITY.
 FT DISULFID 556 573 BY SIMILARITY.
 FT DISULFID 576 585 BY SIMILARITY.
 FT DISULFID 589 610 BY SIMILARITY.
 FT DISULFID 613 621 BY SIMILARITY.
 FT DISULFID 617 629 BY SIMILARITY.
 FT CARBOHYD 126 126 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 250 250 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 353 353 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 408 408 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 414 414 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 437 437 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 469 469 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 522 522 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 566 566 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 616 616 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT VARSPLIC 141 183 ELISGGVYIEKNDKLCMDTIDWRDIVDRDAEIVVKNDR
 FT VARSPLIC 184 1342 VPVTLAAV (IN SHORT ISOFORM).
 FT VARSPLIC 184 1342 MISSING (IN SHORT ISOFORM).
 FT CONFLICT 560 560 E -> G (IN REF. 2).
 FT CONFLICT 1064 1064 E -> G (IN REF. 2).
 FT CONFLICT 1342 1342 AA; 148097 MW; 7201E7F66CA374BD CRC64;
 SQ SEQUENCE 1342 AA; 148097 MW; 7201E7F66CA374BD CRC64;
 Query Match 35.8%; Score 2440.5; DB 1; Length 1342;
 Best Local Similarity 40.7%; Pred. No. 1.8e-122;
 Matches 534; Conservative 191; Mismatches 457; Indels 129; Gaps 32;
 10 GLLALLPQAA--STQVCTGTDMKRLRLPASPTHLDMLRLHYLQGVQVQGNLELYLPT 67
 11 GLLFLARGSEVNSQAVCPGLTNGLSVTDGAENQYQTLKLYERCEVVMGNLEIVLTGH 70
 68 NASLSFLQDIQVQGVYLIHNOVROVPLQRLIRVRGTOLFEDNYALAVLDNGDPLNNTT 127
 71 NADLSFLOWIREVTGVYLVVAMNEFTLPLNLRVVRGTQYDQKFAIFVW-----LNYNT 125
 128 PVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILWKDIPHKNNQALTLIDTN 187
 126 ---NSSHALRQLRLTQLTEILSGGVYIEKNDKLCMDTIDWRDIVDRD---AEIVVKD 178
 188 RSPACHPCPMCKSGSCWSESSDDCSLRTVTCAGGC-ARCKGPLFTDCHEQCAAGCTG 246
 179 NGRSCPPCHEVCKG-RCWPGSDDCQTLTKTICAPQCNHGCFGNPNQCCHECAGCGSG 237
 247 PKHSDCLACLFHNSHICELHCPALVTYNTDTPESMPNDEGRVTFGCACTACPVNVLST 306
 238 PQDTCDFACRHNDSACVPRCPQPLVYNNKLTFLQEPNPHTKYQGVGVASCAPHNFV-V 296
 307 DVGSCTVLCPHLNQVETABDQRCBKSKPCARVCYGLGMEHLREVRVAVTSANIOEFAG 366
 297 DQTSVCVRACPPDRMEVD-KNGLKWCPECGCLCPKACEGTGSG--SRFQVTDSSNIDGFVN 353
 367 CKKIFOSLAFPLPSFGDPSANTAPLOPEOLQVFTLEITGVLYISAWPDSLDPLSVFQ 426
 354 CTKILGNLDFLITGLNGDPWHKIPALDPKLNVRFTVREITGYINTQSWPPHMFNSVFS 413
 427 NLQVIRGRILHNGAYS-LTLQIGIGISWLGRLSRLRELGSGLALIHNTLFCVHTVDPDQL 485
 414 NLTTIGRSLYNGRGSLLIMKNLNVTSGLFRSLUKEISAGRIYISANROLCYHHSNWTKV 473

QY 486 FRNPHQALLHTA-NRPEDBCVGEGLACHOLCARGHCWGPGPTQCVNCVSOFLRGQECVBEC 544
 DB 474 LRGPTEERLDIKHNRPRDCVAEGKVCVDCPLCSGGCGWPGQCLSCRNYSGGVVTHC 533
 QY 545 RVLQGLPREYVNVARHCLPCHPCQPCQNGSVTCFGEADOCVACAHYKDPFCVCARCPGV 604
 DB 534 NFNLEGPREFAEAECSCHPCQWEGTATCNGSGSDTCAOCAHPRDPCVSSCPHGV 593
 QY 605 KPDLSYMPIWKFPDBEGACQPCPINCTHSCVDLDDKGCPEQRA-----SPLTSIVSAVVG 660
 DB 594 LG--AKGPIYKVPDQNECPCHENTCGCKGPELQDCLGQTLVLIGKTHLTWALTVIAG 651
 QY 661 ILLVVVLGVVFGILLIKRQOKIR-KYTMRLLOETELVEPLTPSGAMPNQAOHRIKETE 719
 DB 652 --LVVIFMMLGGTFLYWRGRIQNKRAMRRYLERGESIEPLDPS-EKANKVLARIKETE 708
 QY 720 LRKVKVLGSGFAGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEIIDEAVVMAGVGS 779
 DB 709 LRKLKVLGSGVFTVHKGWIPEGESIKIPVCIKVIEDKSGRQSFQAVTDMHMLAIGSLDH 768
 QY 780 PYVSRLLGICLTSTVQLVTQLMPYGLDHPVRENRLGSGQDLNMCQIAKGMSYLEDV 839
 DB 769 AHVRLGLGCPGSSQLVQYPLGSLLDHVRHQRGALGPQLLNWGVQIAKGMYYLEEH 828
 QY 840 RLVRHDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPPIKMALESILRRRF 899
 DB 829 GNVHRNLARNVLLKSPQVADFGVADLLPDDKQLLYSEAKTPIKMALESIHFGKY 888
 QY 900 THQSDVMSYGVTVWELMTFGAKPDYDIPARETPDLLEKGERLPOPPICITDIVVMVWKW 959
 DB 889 THQSDVMSYGVTVWELMTFGAEFYAGLRLAEPDLEKGERLAQPOICTIDIVVMVWKW 948
 QY 960 MIDSECRPRFRELSEFSRMARDPQRFVVIQNEIDGPA---SPLDSTFYRSLLEDDMDG 1016
 DB 949 MIDENIRPTFKELANEFTEMARDPRLYLVIKES-GPGIAPGEPHGLTNKLEVELEP 1007
 QY 1017 LVDABEYLVPOQGFPCPDPAAGAGMVHRRHSSSTRSGGLDTTGLEP-SSEEA PRSPL 1075
 DB 1008 ELDLDLDEAEED-----NLATTLSALSPLVGTNLNPRGQSLL 1048
 QY 1076 ABSEGAGSDVFDGLGMNAKGLQSLPTH-D-PSPLQRYSEDDTVLP-----SETDGYV 1128
 DB 1049 SPSSGY-MPMNGNIGLSCQESAVSGSSERCPRPSLH-----PMRCLASESSEGHV 1101
 QY 1129 A-----PLTCSQPE-----VYNQDVRPQPPSPREGP-----L 1157
 DB 1102 TGSEAELOEKVSMCRSRSRSPRGDSAYHSQRHSLLTPTVPLSPGLEEDVNGYVM 1161
 QY 1158 PAARFAGATLERAKTLSP-GKNGVV-----KDVFAFGAVENPEYLTPOGGAAPQHPPP 1210
 DB 1162 POTHKGTSPSSREGTSSVGLSSVLGTBEED-----EYEMNNRRRHSP-PHPP 1212
 QY 1211 PAFSFAFDNLYWD-----QDPPERGAPSTFKGTPTAENPEYL 1249
 DB 1213 RPSLEELGEYMDVGSGLSASLGSTQSCPLHPVIMPITAGTTPDEDYEYM 1263
 RESULT 10
 ERB3_RAT
 ID ERB3_RAT STANDARD; PRT; 1339 AA.
 AC Q62799; Q62955;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)
 DE (c-erbB3).
 GN ERB3.
 OS Rattus norvegicus (rat).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN=Sprague-Dawley; TISSUE=Liver;
 RX MEDLINE=96096535; PubMed=8522190;
 RA Hellyer N.J., Kim H.-H., Greaves C.H., Sierke S.L., Koland J.G.;
 RT "Cloning of the rat ErbB3 CDNA and characterization of the
 RL recombinant protein.";
 RL Gene 165:279-284 (1995).
 RN [2]
 RP REVISIONS TO 85; 513 AND 565.
 RA Hellyer N.J., Koland J.G.;
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 922-1097 FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Sciatic nerve;
 RX MEDLINE=97184212; PubMed=9030624;
 RA Carroll S.L., Miller M.L., Frohert P.W., Kim S.S., Corbett J.A.;
 RT "Expression of neuroregulins and their putative receptors, ErbB2 and
 RL ErbB3, is induced during Wallerian degeneration.";
 RL J. Neurosci. 17:1642-1659 (1997).
 CC -1- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTAK.
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
 CC (POTENTIAL).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
 CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
 CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
 CC AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
 CC PHOSPHATIDYLINOSITOL 3-KINASE.
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U29339; AAC28498.2; -;
 DR EMBL; U52530; AAC3050.1; -;
 DR HSSP; P11362; IFGK.
 DR InterPro; IPR000494; EGFR_L_domain.
 DR InterPro; IPR000719; Euk_pkinase.
 DR InterPro; IPR002174; Furin-like.
 DR InterPro; IPR001245; Tyr_pkinase.
 DR Pfam; PF00069; pkinase; 1.
 DR Pfam; PF00757; Furin-like; 1.
 DR PRINTS; PR01030; Recep_L domain; 2.
 DR PRINTS; PR00109; TYRKINASE.
 DR ProDom; PD000001; Euk_pkinase; 1.
 DR SMART; SM00261; FU; 5.
 DR SMART; SM00219; TykC; 1.
 DR PROSITE; PS00107; PROTEIN KINASE ATP; FALSE_NEG.
 DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.
 DR PROSITE; PS00011; PROTEIN KINASE DOM; 1.
 KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
 KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
 FT SIGNAL 1 19
 FT CHAIN 20 1339
 FT DOMAIN 20 643
 FT TRANSMEM 644 662
 FT DOMAIN 663 1339
 FT DOMAIN 183 259
 FT DOMAIN 707 964
 FT NP_BIND 713 721
 FT BINDING 740 740
 FT ACT_SITE 832 832
 FT DISULFID 186 194
 FT DISULFID 190 202
 FT DISULFID 210 218
 FT DISULFID 214 226
 FT DISULFID 227 235

FT DISULFID 231 243 BY SIMILARITY.
 FT DISULFID 246 255 BY SIMILARITY.
 FT DISULFID 259 286 BY SIMILARITY.
 FT DISULFID 290 301 BY SIMILARITY.
 FT DISULFID 305 320 BY SIMILARITY.
 FT DISULFID 323 327 BY SIMILARITY.
 FT DISULFID 500 509 BY SIMILARITY.
 FT DISULFID 504 517 BY SIMILARITY.
 FT DISULFID 520 529 BY SIMILARITY.
 FT DISULFID 533 549 BY SIMILARITY.
 FT DISULFID 556 573 BY SIMILARITY.
 FT DISULFID 576 585 BY SIMILARITY.
 FT DISULFID 589 610 BY SIMILARITY.
 FT DISULFID 613 621 BY SIMILARITY.
 FT DISULFID 617 629 BY SIMILARITY.
 FT CARBOHYD 126 126 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 250 250 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 353 353 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 408 408 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 414 414 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 437 437 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 469 469 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 522 522 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 566 566 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 616 616 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CONFLICT 1028 1028 L -> P (IN REF. 3).
 SQ SEQUENCE 1339 AA; 147545 MW; 0AA5F2402BBFDF1E CRC64;
 Query Match 34.8%; Score 2369.5; DB 1; Length 1339;
 Best Local Similarity 41.0%; Pred. No. 1.1e-118;
 Matches 526; Conservative 171; Mismatches 431; Indels 155; Gaps 34;
 QY 3 LAALCRWGLLLALLPEGAA---STVCTCTDKMLRLPASPTHLDMLRLHYGCGVQVGN 59
 DB 7 LQVLC-----FLSLARGSEMGNQAVCPGTNGLSVTGADNQYQTYLYKLYKEVVMGN 62
 QY 60 LELTLPTNLSLFLQDIQEVQYVLIHINQVQVPLQRLIRIVRGTLQFEDNYALVDN 119
 DB 63 LEIVLTGHNADLSFLQWIREVTGYVLVANVESVLPPLNLRVVRGTQVYDGFAPVM-- 120
 QY 120 GDPLNTTPTVGASPGGLRELQRLSLEILKGGVLIQNPQLCYQDTIILWKDIFHKNNQL 179
 DB 121 ---LNYNT---NSSHALRQLKFTQTLTEILSGGVYIEKNDKLCMDTDIWRDIVRVR--- 170
 QY 180 ALTLDTNRSRACHPCSPMKGSRGWESSEDCOSLTRVTVAGGC-ARCKGPLPTDCCHE 238
 DB 171 GAELVVKNGANGCPPCHEVCKG-RCWGPDPDQCLTKITCAPQCNCRGFCGPNPNCCHD 229
 QY 239 QCAAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTA 298
 DB 230 ECAGGCGSGPQDTCFACRFNDSGACVPRCPPELVYVKLTFTQLEPNPHTKYQYGGVCVAS 289
 QY 299 CPYNYLSTDVGSCTLVCPHINQEVTAEDQTCEKSKFCARVCYGLGMEHLREVRVTS 358
 DB 290 CPNHFV-VDQTFVCRACPPDKMEVD-KHGLKMKCEPCGGLCPKACEGTSGS--SRYQTVD 345
 QY 359 ANIQFAGCKKFGSLAFLESPFDGDPASNTAPLOPEQLQVFTLEETGLYLVISAMPDS 418
 DB 346 SNIDGFVNCKILGNLDFLITGLNVDPMHKIPALDPEKUNVFTVREITGLYLVISAMPDS 405
 QY 419 LPDLSVFQNLQVIRGRILHNGAYS-LTLOGLGISWLGRLSLRELGSGLALIHNTLFCV 477
 DB 406 MNFVSFSLNTTIGRSLYNRGFSLLIMKNLVNLSLGFSLKEISAGRVVYSANQQLCYH 465
 QY 478 HTVPDQLPRNPHQALLHTA-NRPEDECVGEGLAHQLCARGHCWGPGTQVCNCSQFLR 536
 DB 466 HSLNWTLLRGESEERLDIKYDRPLGCELAEGKVCPLCSCSGGCMGPGGCLSCRNYR 525
 QY 537 GOECVEECRVLOGLPREVYNARHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPC 596
 DB 526 EGVCTHNCNLFQGEPEFVHEAQCFSCHEPLMGTSTCNCGSGSDACARCAHFRDGPCH 585
 QY 597 VARCPSGVKPDLSYMPINWKFDPDEGACQPCINCTHSC--VDLDDKGCPEQRASPLTSI 654

Db	609	ISNAYK--FDNRCTCKICHPECR-----TCNGAGADHCQECVHRDQGHCVSCPKNKYN	660
Qy	602	-----SGVK-----PDLSYMPTW	614
Db	661	DGVGCRECHATCGDCTGPKDTIGIGACTTCNLAIINNDATVKRCLLKDDKCPD-GY--FW	717
Qy	615	KF--PBEGACQP-----CPI-----NTH-----	632
Db	718	EYVHPQEGSLKPLAGRAVCRKCHPLCELCTNNGYHEQVCSKCTHYKRRBQCETECPADH	777
Qy	633	-----SC-----VDLDKG-----CPAEQ	646
Db	778	YTDREORECFQRHPENGCTGPGADCKSCRNPKLFDANETGYNVSTWPNCTSKCPLEM	837
Qy	647	R-----ASPLTS-----IVSAVVGILLVVLGVVFGVILIKRRQ	679
Db	838	RHVNYQYTAIGPPYCAASPRSSKITANLDVNMFIITGAVLVFTICILCV--TYICRQK	895
Qy	680	QKIRKYT--MRLQLQETELVEPLTPSGAMPNQAMRILKETELRKVKVLGSGAGFTVYKQ	737
Db	896	QKAKKETVMTWALSGCEDSEPLRPSNIGANLCKLRIVKDAELRKGVLMGAGFGRVYKQ	955
Qy	738	INTPDGENVKIPVAIKVLRNTPSKANKELIDRAYVMAGVSGPYVSRLLIGLICLTSTVQLV	797
Db	956	VWPEGENVKIPVAIKELKSTGAESSEBFLREAYIMASEEHVNLKLLAVCMSSQWMLI	1015
Qy	798	TQLMPYGCLLDHVRENKRGSDOLLNWCQIAKGMGSLYEDVRLVHRDLAARNVLKSPN	857
Db	1016	TQLMPYGCLLDYVNRNDRKIGSKALLNWSQIAKGMGSLYEDVRLVHRDLAARNVLKSPN	1075
Qy	858	HVKITDFGLARLDIDETEHADGGKVPKPKMALESILRRRFTHQSDVMSGYVTVMELMT	917
Db	1076	LVKITDFGLAKLLSSDSNEYKAAGGRWPKWLALECIERNRVFTSKSDVNAFGVTIWEILL	1135
Qy	918	FGAKPYDGIIPAREIPOLLKEGRELPOPPCTIDVYIMVWKCMIDSECRPRRELVSERS	977
Db	1136	FGORPHENIPAKDIPDLIEVGLKLEQPEICSLDIYCTLLSCHWLDAMRPTFKQLTTVEA	1195
Qy	978	RWARDPQRFVWQNEDLG--PASPLDSTFVRSILLEDD--DMGDLVDAEYLVPPQGFPC	1032
Db	1196	EFARDPGRVLAIPGDKFTLPA-----YTSQDEKDLIRKLAPTDGEAKAPDDYILQ	1248
Qy	1033	PDPAAGAGVHRRSSSTRSGGDLTLGLEPSEEEAP-----RSLAPSEAGSDVF	1086
Db	1249	PRAAPGFS-----HRTDCT-----DEMPKLNRYCKDPSKNSSITGDDER	1287
Qy	1087	DG----DLGMAAKGLOSLPHTDPSPLQRYSEDTPVLPSETDGYVAPLTCSPQPEYVNPQ	1143
Db	1288	DSSAREVGVNLR-----LDLPVEDDDYLMF--TCQPGPNNNNM	1325
Qy	1144	DVRPOPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKVDVAFAGAVENPEYL----	1199
Db	1326	N-----NFQNQNMVAAGVAM-----DLGVFVSDNPEYLLNAQTL	1364
Qy	1200	QCGAAPPQH-----PPPAFSP-AFDNLYYWD	1224
Db	1365	GVGESPIPTOTIGIPWVGPGTMEVKVPMFGSGEPTSSDHEYND	1408
RESULT 12			
ERBB	ALV	STANDARD;	PRT; 634 AA.
AC	P00534;		
DT	21-JUL-1986	(Rel. 01, Created)	
DT	15-JUL-1999	(Rel. 38, Last sequence update)	
DT	15-JUN-2002	(Rel. 41, Last annotation update)	
DE	Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).		
GN	V-ERBB.		
OS	Avian leukemia virus.		
OC	Viruses; Retroviral viruses; Retroviridae; Alpharetrovirus.		
OX	NCBI_TaxId=11864;		
RN	[1]		

RP	SEQUENCE FROM N.A.		
RX	MEDLINE=85228222; PubMed=2988784;		
RA	Nilssen T.W., Maroney P.A., Goodwin R.G., Rottman F.M.,		
RA	Crittenden L.B., Raines M.A., Kung H.-J.,		
RT	"c-erbB activation in ALV-induced erythroblastosis: novel RNA		
RT	processing and promoter insertion result in expression of an		
RL	amino-truncated EGF receptor.";		
RL	Cell 41:719-726 (1985).		
CC	-!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein		
CC	tyrosine phosphate.		
CC	-!- MISCELLANEOUS: THIS PROTEIN IS PRODUCED BY ERYTHROLEUKEMIA CELLS		
CC	IN CHICKS WHICH HAVE BEEN INJECTED WITH THE AVIAN LEUKOSIS VIRUS		
CC	AND SUBSEQUENTLY DEVELOP LEUKEMIA. THE LEUKEMIAS ARE INDUCED BY		
CC	THE INSERTION OF RETROVIRAL PROMOTER ELEMENTS IN OR NEAR THE		
CC	C-ERBB GENE, WHICH LEADS TO C-ERBB ACTIVATION.		
CC	-!- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-ENV-ERBB		
CC	PROTEIN.		
CC	-!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL; M10066; AAA48763.1; ALT_INIT.		
DR	PIR; A00643; TVCHLV.		
DR	PIR; B00643; TVFVLV.		
DR	HSSP; P11362; 1FGK.		
DR	InterPro; IPR000719; Euk pkinase.		
DR	InterPro; IPR001245; Tyr_pkinase.		
DR	Pfam; PF00069; pkinase; 1		
DR	PRINTS; PR00109; TYRKINASE.		
DR	ProDom; PD000001; Euk pkinase; 1.		
DR	SMART; SM00219; TyrcK; 1.		
DR	PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.		
DR	PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.		
DR	PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.		
KW	Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;		
KW	Glycoprotein; Phosphorylation.		
FT	DOMAIN	132 399	PROTEIN KINASE.
FT	NP BIND	138 146	ATP (BY SIMILARITY).
FT	BINDING	165 165	ATP (BY SIMILARITY).
FT	ACT SITE	257 257	BY SIMILARITY.
SQ	SEQUENCE	634 AA; 70891 MW; E705E33A0BE01FCC	CRC64;
Query Match 25.7%; Score 1749.5; DB 1; Length 634;			
Best Local Similarity 52.3%; Pred. No. 4.2e-86;			
Matches 370; Conservative 79; Mismatches 135; Indels 123; Gaps 17;			
Qy	587	CAHYKDPFCVACPSGVKPDLSYMPIWKFDPDEGACQPCINCHSCVDLDDKCGPABQ	646
Db	3	CAHFDGPHCVKACPAVGLGENDTL-VMKYADANAVCOLCHENCNTRGCKGPGLEGCP----	58
Qy	647	RASPLTSTVSAV-VGILLVVLGVVGLILIKRQCKIRKYTMRLLOETELVEPLTPSCA	705
Db	59	NGSKTPTSAAGVVGGLCLLVVGLGILYLR--HIVKRTLRLLQRELVEPLTPSGE	117
Qy	706	MPNQAMRILKETELRKVKVLGSGAGFTVYGIWIPDGENVKIPVAIKVLRNTPSKANK	765
Db	118	APNOAHLRLKETETFKVKVVLGSGAGFTVYKGLWIPEGEKVKIPVAIKELRATSPKANK	177
Qy	766	EILDEAYVMAGVSGPYVSRLLIGLICLTSTVQLVQTMYPYGCLLDHVRENKRGSDOLLN	825
Db	178	EILDEAYVMASVDNPHVCRLLIGLICLTSTVQLITLMPYGCLLDIYREHKDNIGSOYLLN	237
Qy	826	CMQIAKMSYLEDVLRDLAARNVLKSNHVKITDFGLARLDIDETEHADGGKVP	885
Db	238	CVQIAKGMNLEERLVRDLAARNVLKTPQHVKITDFGLAKLIGADEKEYHAEKGGKVP	297
Qy	886	IKWMALESILRRRFTHQSDVMSGYVTVMELTFGAKPYDGIIPAREIPDLEKGERLPQP	945


```
Db 298 IKWMALESILHRIYTHOSDVSYGVTWELMTFGSKPYDGPASEISSVLEKGERLPQPP 357
Qy 946 ICTIDVYIMVKCWMIDSECRPFRELVSFSESRWARDPQRFVVIQ-NEDLGPASPLDSTF 1004
Db 358 ICTIDVYIMVKCWMIDADSRPFRELIAEFSKWARDPPRYLVIQGDERMHLPSPTDSKF 417
Qy 1005 YRSILEDMDGLVDABEYLVPOQGFPCDPAPGAGGMVHRRSSSTRSGGDLTLGLE 1064
Db 418 YRTLMEEDMEDIVDAEYLVPHQGF-----NSPST----- 449
Qy 1065 PSEEEAPRSP-----APSEGAGSDVDFDGLGMAAKGLQSLPTHDPSPLOQRYSEDP 1119
Db 450 -----SRTPLLSLSLTSATNSATNCID-----RNGQGHVPREDSFVQRYSSDPTGN 495
Qy 1120 LPSET--DGIVAPLTCSPQPEYUNQDVRPQPSREGPLPAARPAGATILERAKTLS 1177
Db 496 FLEESIDDDGLF-----PAPEYVQ--LMPKKFS-----TAMVQ 526
Qy 1178 NGVVKDVF-----AFGAVENPEYLTPOGGAAPQPHPPAFSPAFONL 1222
Db 527 NQIYNNISLTAISKLPMSRYQNSHSTAVDNPEYL-----NTNQSPLAKTVFESSPY 578
Qy 1223 WDQ-----DPPE-----RGAPSTFKGTPTAENPEYLGNDVP 1254
Db 579 WIOGNGHQLNDPNFYQDQLPNETKNGLLKVPAAENPEYLRVAAP 625

RESULT 13
ERBB AVIER
ID ERBB AVIER STANDARD; PRT; 604 AA.
AC P00535;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 13-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN V-ERBB.
OS Avian erythroblastosis virus (strain ES4).
OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=79685;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H;
RX MEDLINE=84026539; PubMed=6313229;
RA Yamamoto T., Nishida T., Miyajima N., Kawai S., Ooi T., Toyoshima K.;
RT "The erbB gene of avian erythroblastosis virus is a member of the src
RT gene family.";
RL Cell 35:71-78(1983).
RP [2]
RP SEQUENCE OF 1-152 FROM N.A.
RX MEDLINE=84223957; PubMed=6328658;
RA Debuire B., Henry C., Benaisa M., Biserte G., Claverie J.-M.,
RA Saule S., Martin P., Stehelin D.;
RT "Sequencing the erbA gene of avian erythroblastosis virus reveals a
RT new type of oncogene.";
RL Science 224:1456-1459(1984).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- DISEASE: THE V-ERBB ONCOGENE TRANSFORMS AVIAN FIBROBLASTS AND
CC ERYTHROBLASTS IN CULTURE AND INDUCES SARCOMAS AND ERYTHROLEUKEMIAS
CC IN CHICKENS.
CC -1- MISCELLANEOUS: V-ERBB IS A TRUNCATED AND MUTATED VERSION OF THE
CC RECEPTOR FOR EPIDERMAL GROWTH FACTOR.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
```

```
DR EMBL; K02006; AAA42394.1; ALT_INIT.
DR EMBL; K01216; AAA42400.1; -.
DR PIR; A00644; TVYUH.
DR HSSP; P11362; LFQK.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR Transfaser; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW Glycoprotein; Phosphorylation.
FT DOMAIN 132 399 PROTEIN KINASE.
FT NP_BIND 138 146 ATP (BY SIMILARITY).
FT BINDING 165 165 ATP (BY SIMILARITY).
FT ACT_SITE 257 257 BY SIMILARITY.
FT CONFLICT 29 29 R -> W (IN REF. 2).
FT CONFLICT 140 140 S -> F (IN REF. 2).
FT CONFLICT 146 146 I -> V (IN REF. 2).
SQ SEQUENCE 604 AA; 7633 MW; 76BECDD06745D609 CRC64;

Query Match 25.0%; Score 1703; DB 1; Length 604;
Best Local Similarity 52.2%; Pred. No. 1.2e-83;
Matches 360; Conservative 76; Mismatches 128; Indels 126; Gaps 16;

Qy 587 CAHYKDPPECVACPSGVKPDLSYPIWKFPDEEGACQPCINCTHSCVDLDDKGGPAEQ 646
Db 3 CAHFDGPHCVKACPAVLGENDTL--VRKYADANAVCQLCHPNCTRGCKGPGLEGCP--- 58
Qy 647 RASPLTSIVSAVV-GILLVVLGVVFGILIKERQKIRKYTWRLRLQETELVEPLTPSGA 705
Db 59 NSKTPSIAAGVVGGLLVGLGGLYLR--HIVKRTLRLLQERLVEPLTPSGE 117
Qy 706 MPNQAQMRILKETELRKVKVLGSGAPGVYKGIWIPDGENVKIPVAIKVLRNTPSKANK 765
Db 118 APNQAHLRLKETEFKKVKVLGSGAPGVYKGIWIPDGENVKIPVAIKVLRNTPSKANK 177
Qy 766 EILDEAYVWAGVSPVSRLLGLCTSTVOLTPMPYGLLDHVRNRRGLSGQDLLANW 825
Db 178 EILDEAYVWASVDNPHCVCLGLCTSTVOLTPMPYGLLDHVRNRRGLSGQDLLANW 237
Qy 826 CMQIAGMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETVHADGKVP 885
Db 238 CVQIAGMNYLERRLVHRDLAARNVLVKTPOHKITDFGLAKLLGADEKEYHAEGKVP 297
Qy 886 IKWMALESILHRIYTHOSDVSYGVTWELMTFGSKPYDGPASEISSVLEKGERLPQPP 945
Db 298 IKWMALESILHRIYTHOSDVSYGVTWELMTFGSKPYDGPASEISSVLEKGERLPQPP 357
Qy 946 ICTIDVYIMVKCWMIDSECRPFRELVSFSESRWARDPQRFVVIQ-NEDLGPASPLDSTF 1004
Db 358 ICTIDVYIMVKCWMIDADSRPFRELIAEFSKWARDPPRYLVIQGDERMHLPSPTDSKF 417
Qy 1005 YRSILEDMDGLVDABEYLVPOQGFPCDPAPGAGGMVHRRSSSTRSGGDLTLGLE 1064
Db 418 YRTLMEEDMEDIVDAEYLVPHQGF-----NSPST----- 449
Qy 1065 PSEEEAPRSP-----APSEGAGSDVDFDGLGMAAKGLQSLPTHDPSPLOQRYSEDP 1119
Db 450 -----SRTPLLSLSLTSATNSATNCID-----RNGQGHVPREDSFVQRYSSDPTGN 495
Qy 1120 LPSET--DGIVAPLTCSPQPEYUNQDVRPQPSREGPLPAARPAGATILERAKTLS 1177
Db 496 FLEESIDDDGLF-----PAPEYVQ--LMPKKPSTAM----- 524
Qy 1178 NGVVKDVF-----AFGAVENPEYLTPOGGAAPQPHPPAFSPAFD 1218
Db 525 --VQNIYNNISLTAISKLPMSRYQNSHSTAVDNPEYL-----NTNQSPLAKTVFE 574
Qy 1219 NLYYVQDQPPPERGAPSTFKGTPTAENPEY 1248
```



```
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M20386; AAA48760.1; -.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR SMART; SM00261; FU; 4.
DR PROSITE; PS00107; PROTEIN KINASE ATP; PARTIAL.
DR PROSITE; PS00109; PROTEIN KINASE TYR; PARTIAL.
DR PROSITE; PS00011; PROTEIN KINASE DOM; PARTIAL.
DR Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
KW Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT SIGNAL 1 30
FT CHAIN 31 >703 EPIDERMAL GROWTH FACTOR RECEPTOR.
FT DOMAIN 31 654 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 655 667 POTENTIAL.
FT DOMAIN 668 >703 CYTOPLASMIC (POTENTIAL).
FT DISULFID 197 206 BY SIMILARITY.
FT DISULFID 201 214 BY SIMILARITY.
FT DISULFID 222 230 BY SIMILARITY.
FT DISULFID 226 238 BY SIMILARITY.
FT DISULFID 239 247 BY SIMILARITY.
FT DISULFID 243 255 BY SIMILARITY.
FT DISULFID 258 267 BY SIMILARITY.
FT DISULFID 271 298 BY SIMILARITY.
FT DISULFID 302 314 BY SIMILARITY.
FT DISULFID 318 333 BY SIMILARITY.
FT DISULFID 336 340 BY SIMILARITY.
FT DISULFID 513 522 BY SIMILARITY.
FT DISULFID 517 530 BY SIMILARITY.
FT DISULFID 533 542 BY SIMILARITY.
FT DISULFID 546 562 BY SIMILARITY.
FT DISULFID 565 581 BY SIMILARITY.
FT DISULFID 569 589 BY SIMILARITY.
FT DISULFID 592 601 BY SIMILARITY.
FT DISULFID 605 627 BY SIMILARITY.
FT DISULFID 630 638 BY SIMILARITY.
FT DISULFID 634 646 BY SIMILARITY.
FT CARBOHYD 134 134 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 190 190 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 200 200 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 359 359 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 368 368 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 420 420 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 573 573 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 578 578 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 613 613 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 633 633 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 648 648 N-LINKED (GLCNAC. .) (POTENTIAL).
FT NON_TER 703 703
SQ SEQUENCE 703 AA; 77427 MW; AFF2DE11B735A690 CRC64;

Query Match 23.8%; Score 1624; DB 1; Length 703;
Best Local Similarity 44.7%; Pred No 2.2e-79;
Matches 316; Conservative 113; Mismatches 250; Indels 28; Gaps 12;

QY 8 RWGLLALLPPGAA-----STVCTGTDMLRLPASPTHLDMLRHLXGCGVQGNLE 61
DB 13 RGAALVLLLLGVALCSAVEEKKVCCQTNNKLTQLGHVEDHFTSLQRMYNCEVLNLE 72

QY 62 LTYLPTNASLFLQDIOEVGYVLIHNVQVPLQRLRIVRGTLQFEDNYALAVLDNGD 121
DB 73 ITVEHNRDLTFLKTIQEVAGYVLIHNMVDVTPLENLQIRGNVLYDNSFALAVLSNH 132

QY 122 PLNNTTPTVTGASPGGLRELQRLSITELKGGVLIQRPOLCYODTILWKDI FHKNNQAL 181
DB 133 -WNKTO-----GLRELPMKRLSELINGVKISNNPKLNMMDTVLWMDIITSRK-PL 182

QY 182-TLID-TNRSRAPHCSPMKSGSRCWGESSEDQSLTRTVCAAGCA-RCKGPLPTDCHEQ 239
DB 182-TLID-TNRSRAPHCSPMKSGSRCWGESSEDQSLTRTVCAAGCA-RCKGPLPTDCHEQ 239
```

```
DB 183 TVLDFASNLSSCPKHPNCTEDHCWAGEQNQOTLTVKVCAQCSGCRGKVPKPSDCCHNQ 242
QY 240 CAAGCTGPKHSCLACLHFNHSGICELHCPALVYNTDTFESMPNPEGRTVTCASCVTAC 299
DB 243 CAAGCTGPRESDCLACRKFDDATCDDTCTPPLVLYNFTTYQMDVNPBGKYSFGATCVREC 302
QY 300 PYNLSTDVGSCTLVCPHLNQEVTAADGTQRCBKCSKPCARVCYGLGMEHLREVRAVTS 359
DB 303 PHNVVTDHSGSCVRSCTDTYEV-ENGVRKCKCDGLCSKVNGIGIGELKGILINAT 361
QY 360 NIOEFAGCKKIFGSLAFPESDGDPASNTAPLOPEQLQVFEITGLYIYSAMPDSL 419
DB 362 NIDSFKNCTKINGDVSILPVAFLGDAFTKTLPLDPKKLDVFTVKEISGFLLIQAWPDNA 421
QY 420 PDLSPONLOVIRGRILHNGAYSLTLQGLGISWGLGRSLRELGSGLALIHNNTHLCFVHT 479
DB 422 TDLAYAFENLEIRGRTKQHQVSLAVVNLKIOSLGRSLRKEISDSDIAIMKNKNLCYADT 481
QY 480 VPWDQLFRNPHQALLHTANRPEDECYGEGLACHQLCARGHCWGPPTQCVNCSQFLRGOE 539
DB 482 MNWRSLFATQSQKTKIIQNRNKNDCTADRHVCDPLCSDVGCWGPFPFHCFSRQKE 541
QY 540 CVEECRVLQGLPREYVNAHCLPCHPECPQNG---SVTCFGEADQCACAHYKDPFPC 596
DB 542 CVKQCNILQGEPEFERDSKCLPCHSECLVQNSYNTTCSGPGPDHCKMCAHFIDGPHC 601
QY 597 VARCPGVKPDLSYMPIWKFPEDEGACQPCINCTHSCVDLDDKGCPSAQRASPLTSIVS 656
DB 602 VRACPAGVLENDTL-VWKYADANAVCOLCHPNCTRGCKPGLEGCP---NGSKTPSIAA 657
QY 657 AVV-GILLVVVLGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTP 702
DB 658 GVVGGLLVVVLGVVGLGILYLRRL-HIVRKRTLRRLQERELVEPLTP 703
```

Search completed: July 22, 2003, 08:06:37

Job time : 44.0519 secs

Qy	1081	AGSDVFDGLGMGAAGLQSLPTHDPSPLORYSEDPTVPLPSETDGYVAPLTCSPQPEYV	1140
Db	1081	AGSDVFDGLGMGAAGLQSLPTHDPSPLORYSEDPTVPLPSETDGYVAPLTCSPQPEYV	1140
Qy	1141	NQPDVFPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKOVFAFGGAVENPEYLTPO	1200
Db	1141	NQPDVFPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKOVFAFGGAVENPEYLTPO	1200
Qy	1201	GGAAPQHPPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV	1255
Db	1201	GGAAPQHPPPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV	1255

Search completed: July 22, 2003, 08:08:14
Job time : 97.3788 secs

THIS PAGE BLANK (USPTO)

QY 781 YVSRLLGICLTSTVQLTQMPYGLDHDVRENRGRIGLSQDLNMCQIAKMSYLEDDR 840
Db 781 YVSRLLGICLTSTVQLTQMPYGLDHDVRENRGRIGLSQDLNMCQIAKMSYLEDDR 840
QY 841 LVHRDLAARNVLKSPNHVKITDGLARLLDIDETEYHAGGKVPKIMWMALESILRRFT 900
Db 841 LVHRDLAARNVLKSPNHVKITDGLARLLDIDETEYHAGGKVPKIMWMALESILRRFT 900
QY 901 HQSDVMSYGYVWELMTFGAKPDGIPAREIPDLLEKGERLPQPICTIDVYIMVVKCWM 960
Db 901 HQSDVMSYGYVWELMTFGAKPDGIPAREIPDLLEKGERLPQPICTIDVYIMVVKCWM 960
QY 961 IDSECRPRFRELVSFMRMARDQRFVVIQNEDLGPASPLDSTFYRSLLEDMDGLVDA 1020
Db 961 IDSECRPRFRELVSFMRMARDQRFVVIQNEDLGPASPLDSTFYRSLLEDMDGLVDA 1020
QY 1021 EBYLVPOQGFCDPAPGAGVHHRSSSTRSGGDLTLGLEPSEEEAPRSPAPSEG 1080
Db 1021 EBYLVPOQGFCDPAPGAGVHHRSSSTRSGGDLTLGLEPSEEEAPRSPAPSEG 1080
QY 1081 AGSDVFDGDLGMAAGKLSLTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGDLGMAAGKLSLTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
QY 1141 NQPDVRRPQPSREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAPGGAVENPEYLTPO 1200
Db 1141 NQPDVRRPQPSREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAPGGAVENPEYLTPO 1200
QY 1201 GGAAPQHPHPPAPSPAFDNLVYWDQDPPRGAPPSTFKGTPTAENPEYGLDVPV 1255
Db 1201 GGAAPQHPHPPAPSPAFDNLVYWDQDPPRGAPPSTFKGTPTAENPEYGLDVPV 1255

RESULT 14
AAU77114
ID AAU77114 standard; Protein; 1255 AA.
XX AAU77114;
AC XX
XX XX
DT 05-JUN-2002 (first entry)
XX Human Her-2/neu polypeptide.
DE XX
XX Human; Her-2/neu; cytostatic; haematological malignancy; CML;
KW acute myelogenous leukaemia; AML; chronic myelogenous leukaemia; CLL;
KW chronic lymphocytic leukaemia; myeloma; non-Hodgkin's lymphoma; MDS;
KW Hodgkin's lymphoma; T cell therapy.
XX
OS Homo sapiens.
XX
XX WO200213847-A2.
XX
XX 21-FEB-2002.
XX
XX 13-AUG-2001; 2001WO-US25408.
XX
XX 14-AUG-2000; 2000US-0638280.
XX 28-SEP-2000; 2000US-0675904.
XX
XX (CORI-) CORIXA CORP.
XX
XX Gaiger A, Cheever MA, Hand-zimmermann S;
XX
XX WPI: 2002-280741/32.
XX N-PSDB; ABK10730.
XX
XX Inhibiting haematological malignancy development by administering
PT polypeptide comprising immunogenic portion of Her-2/neu, polynucleotide
PT encoding the polypeptide, or antigen presenting cells expressing the
PT polypeptide
XX
XX Disclosure; Page 71-74; 74pp; English.
PS
XX

CC The invention relates to a method for inhibiting development of
CC haematological malignancy in a patient by administering a polypeptide
CC comprising an immunogenic portion of Her-2/neu or a polynucleotide
CC encoding the polypeptide. Antigen presenting cells that express the
CC protein can also be administered. The sequences are used for inhibiting
CC development of haematological malignancy such as acute myelogenous
CC leukaemia (AML), chronic myelogenous leukaemia (CML), chronic lymphocytic
CC leukaemia (CLL), MDS, myelomas, Hodgkin's lymphoma and non-Hodgkin's
CC lymphoma. This sequence represents the human Her-2/neu polypeptide.
XX
SQ Sequence 1255 AA;
Query Match 99.9%; Score 6806; DB 23; Length 1255;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1253; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 MELAAALCRGLLLALLPPGAASTQVCTGDMKRLPASPTHLDMLRHLYQGCVQVGNL 60
Db 1 MELAAALCRGLLLALLPPGAASTQVCTGDMKRLPASPTHLDMLRHLYQGCVQVGNL 60
QY 61 ELYLPTNASLSPLQDIQEVQGVLIHNRQVQVPLQRLRIVRGTLQFEDNYALAVLDNG 120
Db 61 ELYLPTNASLSPLQDIQEVQGVLIHNRQVQVPLQRLRIVRGTLQFEDNYALAVLDNG 120
QY 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIORNPOLCYQDTILMKDIFHNKOLA 180
Db 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIORNPOLCYQDTILMKDIFHNKOLA 180
QY 181 LTLIDTNRSRACHPCSPMCKSGRCWGESSEDCOSLTRTVCAGGCARCKGLPTDCCHEQC 240
Db 181 LTLIDTNRSRACHPCSPMCKSGRCWGESSEDCOSLTRTVCAGGCARCKGLPTDCCHEQC 240
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPGRYTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPGRYTFGASCVTACP 300
QY 301 YNYLSTDVGSCTLVCPHNRQVTAEDGTQRCCKSPCARVCYGLGWEHLREVRVTSAN 360
Db 301 YNYLSTDVGSCTLVCPHNRQVTAEDGTQRCCKSPCARVCYGLGWEHLREVRVTSAN 360
QY 361 IQEFACKKI FGSIAFLPESFDGDPASNTAPLOPEQLQVFETLEEITGYLIYISAWPDSL 420
Db 361 IQEFACKKI FGSIAFLPESFDGDPASNTAPLOPEQLQVFETLEEITGYLIYISAWPDSL 420
QY 421 DLSVFQNLQVIRGRILHNGAYSILTLQGLGSIWGLRSLRSLGSLALIHNTILCFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYSILTLQGLGSIWGLRSLRSLGSLALIHNTILCFVHTV 480
QY 481 PMDQFRNPQALLHTANRPEDECVGEGLAGHLCARHGWGPGTQCVCNCSQFLRGQEC 540
Db 481 PMDQFRNPQALLHTANRPEDECVGEGLAGHLCARHGWGPGTQCVCNCSQFLRGQEC 540
QY 541 VEECRVLQGLPREYVNRHCLPCHPECPQNGSVTCFGEADOCVACAHYKDPFCVARC 600
Db 541 VEECRVLQGLPREYVNRHCLPCHPECPQNGSVTCFGEADOCVACAHYKDPFCVARC 600
QY 601 PSGVKPDLVMPITWKFPEDEGACQPCINCTHSCVDLDDKCPAEORASPLTSIVSAVVG 660
Db 601 PSGVKPDLVMPITWKFPEDEGACQPCINCTHSCVDLDDKCPAEORASPLTSIVSAVVG 660
QY 661 ILLVVVLGVVFGILIKRQOKIRKYTMRLLOETELVEPLTPSGAMPNQAMRILKETEL 720
Db 661 ILLVVVLGVVFGILIKRQOKIRKYTMRLLOETELVEPLTPSGAMPNQAMRILKETEL 720
QY 721 RKVKVLGSGAFGTVYKGIWIPGENVKI PVAIKVLRENTS PKANKIILDEAYVWAGVSP 780
Db 721 RKVKVLGSGAFGTVYKGIWIPGENVKI PVAIKVLRENTS PKANKIILDEAYVWAGVSP 780
QY 781 YVSRLLGICLTSTVQLTQMPYGLDHDVRENRGRIGLSQDLNMCQIAKMSYLEDDR 840
Db 781 YVSRLLGICLTSTVQLTQMPYGLDHDVRENRGRIGLSQDLNMCQIAKMSYLEDDR 840
QY 841 LVHRDLAARNVLKSPNHVKITDGLARLLDIDETEYHAGGKVPKIMWMALESILRRFT 900

Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWKMALESILRRFT 900
Qy 901 HQSDVWSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPPOPICTIDVYIMVKCWM 960
Db 901 HQSDVWSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPPOPICTIDVYIMVKCWM 960
Qy 961 IDSECRPRFRELVSERFARMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDGMDLVD 1020
Db 961 IDSECRPRFRELVSERFARMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDGMDLVD 1020
Qy 1021 EYLVPOQGFCCPDAPGAGVHHRSSSTRSGGDLTLGLEPSEEPAPRPLAPSEG 1080
Db 1021 EYLVPOQGFCCPDAPGAGVHHRSSSTRSGGDLTLGLEPSEEPAPRPLAPSEG 1080
Qy 1081 AGSDVFDGLGMAAGLQSLPHTDPSPIQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGLGMAAGLQSLPHTDPSPIQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Qy 1141 NQPDVRPQPPSPREGPLPAARPAATLERAKTILSPGKNGVVKDVAFGGAVENPEYLTPO 1200
Db 1141 NQPDVRPQPPSPREGPLPAARPAATLERAKTILSPGKNGVVKDVAFGGAVENPEYLTPO 1200
Qy 1201 GGAAPQHPHPPAFSPADNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGIDVPV 1255
Db 1201 GGAAPQHPHPPAFSPADNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGIDVPV 1255

RESULT 15

AAR39568
ID AAR39568 standard; Protein; 1433 AA.

XX AAR39568;

DT 07-FEB-1994 (first entry)

XX Sequence of c-erbB-2 tumour antigen.

XX Tumour antigen; c-erbB-2; glycoprotein.

XX Homo sapiens.

XX WO9316185-A.

XX 19-AUG-1993.

PF 05-FEB-1993; 93WO-US01055.

PR 06-FEB-1992; 92US-0831967.

XX (CETU) CETUS ONCOLOGY CORP.

PA (CREA-) CREATIVE BIOMOLECULES INC.

XX Houston LL, Huston JS, Oppermann H, Ring DB;

XX WPI; 1993-272889/34.

DR N-PSDB; AAQ46083.

XX New single chain Fv polypeptide binding to C-erbB-2 tumour antigen - for imaging or treating breast or ovarian cancer etc.

XX Disclosure; pages 48-54; 87pp; English.

CC c-erbB-2 refers to a protein antigen expressed on the surface of tumour cells, such as breast and ovarian tumour cells, which is an approx. 200,000 mol. wt. acidic glycoprotein having an isoelectric pt. of about 5.3 (see AAQ46083, AAR39568). X in AAR39583 represents the location of a stop codon in AAQ46083.

XX Sequence 1433 AA;

Query Match 99.3%; Score 6763; DB 14; Length 1433;

Best Local Similarity 99.3%; Pred. No. 0;

		Matches 1246; Conservative 3; Mismatches 6; Indels 0; Gaps 0;	
Qy	1	MELAAALCRGGLLALLPPGAASSTQVCTGTDMLRLPASPTHLDMLRHLHYQGVVQGNL	60
Db	1	MELAAALCRGGLLALLPPGAASSTQVCTGTDMLRLPASPTHLDMLRHLHYQGVVQGNL	60
Qy	61	ELTYLPTNASLSFLQDIOEVQGVYLIAHQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG	120
Db	61	ELTYLPTNASLSFLQDIOEVQGVYLIAHQVRQVPLQRLRIVRGTQLFEDNYALAVLDNG	120
Qy	121	DPLNNTPTVGTASPGGLRELQRLSLTEILKGGVLIORNPOLCYQDTILWKDIFHKQNOLA	180
Db	121	DPLNNTPTVGTASPGGLRELQRLSLTEILKGGVLIORNPOLCYQDTILWKDIFHKQNOLA	180
Qy	181	LTLIDNRRGRACHPCSPCKGRCWGESSEDCOSLRTVCAGGCARCKGPLPTDCCHEQC	240
Db	181	LTLIDNRRGRACHPCSPCKGRCWGESSEDCOSLRTVCAGGCARCKGPLPTDCCHEQC	240
Qy	241	AAGCTGPKHSDCLACILHFNHSGICEILHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP	300
Db	241	AAGCTGPKHSDCLACILHFNHSGICEILHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP	300
Qy	301	YNYLSTDVSGSCTLVCPHLNQEVTAEDGTQRCCKSPCARVCYGLGMEHLREVRVTSAN	360
Db	301	YNYLSTDVSGSCTLVCPHLNQEVTAEDGTQRCCKSPCARVCYGLGMEHLREVRVTSAN	360
Qy	361	IQEFAGCKIFGSLAFLPESFDGDPASNTAPLOPELOVPFETLEITGVLYISAWPDSL	420
Db	361	IQEFAGCKIFGSLAFLPESFDGDPASNTAPLOPELOVPFETLEITGVLYISAWPDSL	420
Qy	421	DLVSFQNLQVIRGRILHNGAYSLTLQGLGSIWGLRSLRSLRSLRSLRSLRSLRSLRSLR	480
Db	421	DLVSFQNLQVIRGRILHNGAYSLTLQGLGSIWGLRSLRSLRSLRSLRSLRSLRSLRSLR	480
Qy	481	PMDQLFRNPHQALLHTANRPEDECVGEGACHQLCARGHGWGPGTQCVNCSOFLRGQEC	540
Db	481	PMDQLFRNPHQALLHTANRPEDECVGEGACHQLCARGHGWGPGTQCVNCSOFLRGQEC	540
Qy	541	VVECRVLQGLPREYVNAHCLCHPECPQNGSVTCFGEADOCVACAHYKDPFCVARC	600
Db	541	VVECRVLQGLPREYVNAHCLCHPECPQNGSVTCFGEADOCVACAHYKDPFCVARC	600
Qy	601	PSGVKPDLSYMPIWKPPEDEGACQPCINCTHSCVDLDDKGPAPQASPLTSIVSAVVG	660
Db	601	PSGVKPDLSYMPIWKPPEDEGACQPCINCTHSCVDLDDKGPAPQASPLTSIVSAVVG	660
Qy	661	ILLVVVLGVVFGILIKRROOKIRKYTMRLLOETELVEPLTPSGAMPNQAQRILKETEL	720
Db	661	ILLVVVLGVVFGILIKRROOKIRKYTMRLLOETELVEPLTPSGAMPNQAQRILKETEL	720
Qy	721	RKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKELDEAYVMAGVSP	780
Db	721	RKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKELDEAYVMAGVSP	780
Qy	781	YVSRLLIGICLTSTVQLVTLQMPYGCLLDHVRENRLGSDLLNWCWQAKGMSYLEDDR	840
Db	781	YVSRLLIGICLTSTVQLVTLQMPYGCLLDHVRENRLGSDLLNWCWQAKGMSYLEDDR	840
Qy	841	LVRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWKMALESILRRFT	900
Db	841	LVRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWKMALESILRRFT	900
Qy	901	HQSDVWSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPPOPICTIDVYIMVKCWM	960
Db	901	HQSDVWSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPPOPICTIDVYIMVKCWM	960
Qy	961	IDSECRPRFRELVSERFARMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDGMDLVD	1020
Db	961	IDSECRPRFRELVSERFARMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDGMDLVD	1020
Qy	1021	EYLVPOQGFCCPDAPGAGVHHRSSSTRSGGDLTLGLEPSEEPAPRPLAPSEG	1080
Db	1021	EYLVPOQGFCCPDAPGAGVHHRSSSTRSGGDLTLGLEPSEEPAPRPLAPSEG	1080

Db 1021 EBYLVPOQGFCDPAPGAGMVHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080
Qy 1081 AGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVVAPLTCSPOPEYV 1140
Db 1081 AGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVVAPLTCSPOPEYV 1140
Qy 1141 NQPDVRPQPPSPREGPLPAARAGATLERAKTILSPGKNGVVKOVFAFGGAVENPEYLTPO 1200
Db 1141 NQPDVRPQPPSPREGPLPAARAGATLERAKTILSPGKNGVVKOVFAFGGAVENPEYLTPO 1200
Qy 1201 GGAAPQHPHPPAPSPAFDNLVYWDQDPPERGAPSPFTKGTPTAENPEYLGIDVVP 1255
Db 1201 GGAAPQHPHPPAPSPAFDNLVYWDQDPPERGAPSPFTKGTPTAENPEYLGIDVVP 1255

RESULT 13

AA051143
ID AA051143 standard; Protein; 1255 AA.

XX AA051143;
XX 17-JUN-2002 (first entry)
XX Human Her-2/neu oncogene-encoded p185 glycoprotein.
XX Her-2/neu; oncogene; cancer; tumour; vaccine; human; p185;
KW tyrosine kinase; receptor; c-erbB2; gene therapy.
XX Homo sapiens.

XX Key Location/Qualifiers
FH Domain 1..653 /note= "extracellular domain"
FT Domain 676..1255 /note= "intracellular domain"
FT Domain 990..1255 /note= "phosphorylation domain"

XX W0200212341-A2.
XX 14-FEB-2002.
XX 03-AUG-2001; 2001WO-US24283.
XX 03-AUG-2000; 2000US-0632507.
XX (CORI-) CORIXA CORP.
XX (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

XX Cheever MA, Gheysen D;
XX WPI; 2002-241743/29.
XX N-PSDB; ABA92250.
XX Her-2/neu fusion protein for treating or preventing cancer by eliciting
PT or enhancing an immune response to the protein, has Her-2/neu
PT extracellular domain fused to Her-2/neu intracellular or
PT phosphorylation domain
XX Claim 68; Fig 7; 141pp; English.

XX The present sequence is that of human Her-2/neu (p185 glycoprotein
CC or c-erbB2), an oncogenic self-protein and target for anti-cancer
CC vaccines. The Her-2/neu gene is amplified and p185 is overexpressed
CC in a variety of cancers, including breast, ovarian, colon, lung and
CC prostate cancer. Her-2/neu is a member of the tyrosine kinase
CC family of receptor-like glycoproteins. It comprises an extracellular
CC domain with homology to the epidermal growth factor receptor
CC (EGFR), a highly hydrophobic transmembrane domain and a C-terminal
CC intracellular domain that also shows homology to EGFR. Its
CC overexpression correlates with a poor prognosis in breast and
CC ovarian cancers. The invention provides Her-2/neu fusion
CC proteins, nucleic acids encoding them, viral vectors, and vaccines

CC comprising the fusion proteins or nucleic acid molecules. In
CC preferred fusion proteins, the extracellular domain of a Her-2/neu
CC protein is fused to a Her-2/neu intracellular domain or
CC phosphorylation domain (or its DeltaPP fragment). An immune
CC response to Her-2/neu protein is elicited or enhanced by
CC administering the fusion protein in the form of a vaccine, or by
CC transfusing cells of an animal ex vivo with a nucleic acid
CC encoding the fusion protein, and delivering the transfected cells
CC to the animal. The fusion proteins, nucleic acids, and isolated
CC specific T-cells are useful for inhibiting the development of a
CC cancer, especially breast, ovarian, colon, lung or prostate cancer
CC in a patient. T cells that specifically react with a Her-2/neu
CC fusion protein can be used to remove tumour cells from a sample in
CC order to inhibit the development of cancer in a patient.

XX SQ Sequence 1255 AA;

Query Match 99.9%; Score 6806; DB 23; Length 1255;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1253; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MELAALCRWGLLLALLPPGAASQVCTGTDMLRLPASPTHLDMLRHLHYQGCVVQGNL 60
Db 1 MELAALCRWGLLLALLPPGAASQVCTGTDMLRLPASPTHLDMLRHLHYQGCVVQGNL 60
Qy 61 ELTYLPTNASLSFLQDIEQVGVYLIHQNVRQVPLQRLRIVRGSTQIFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIEQVGVYLIHQNVRQVPLQRLRIVRGSTQIFEDNYALAVLDNG 120
Qy 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQRPOLCYQDTILWKDIHFNKNOLA 180
Db 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQRPOLCYQDTILWKDIHFNKNOLA 180
Qy 181 LTLIDTNRSRACHPCSPMCKGSRGCESSDCOSLRTVCAGCARKGKGLPTDCCHEQC 240
Db 181 LTLIDTNRSRACHPCSPMCKGSRGCESSDCOSLRTVCAGCARKGKGLPTDCCHEQC 240
Qy 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Qy 301 YNYLSTDVSGCTLVCPHNOEVTAEADGTQRCCKSPCARVCYGLGMEHLREVRVTSAN 360
Db 301 YNYLSTDVSGCTLVCPHNOEVTAEADGTQRCCKSPCARVCYGLGMEHLREVRVTSAN 360
Qy 361 IOEFAGCKITFGSLAFLPESFDGPASNTAPLOPEQLQVFTELEETGYLYISAWPDSL 420
Db 361 IOEFAGCKITFGSLAFLPESFDGPASNTAPLOPEQLQVFTELEETGYLYISAWPDSL 420
Qy 421 DLSVFQNLQVIRGRILHNGAYSLTLOGLISWLGRLSRLRELGSGLALIHNTLCFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYSLTLOGLISWLGRLSRLRELGSGLALIHNTLCFVHTV 480
Qy 481 PWDOLFRNPHQALLHTANRPEDECVGEGACHOLCARGHCWGPGTQCVNCSQFLRGQEC 540
Db 481 PWDOLFRNPHQALLHTANRPEDECVGEGACHOLCARGHCWGPGTQCVNCSQFLRGQEC 540
Qy 541 VEECRVLQGLPREYVVARHCLPCHPECPQNGSVTCFGEADOCVACAHYKDPFCVARC 600
Db 541 VEECRVLQGLPREYVVARHCLPCHPECPQNGSVTCFGEADOCVACAHYKDPFCVARC 600
Qy 601 PSQVKPDLSPYMPWKFPDEBEGACQPCINCTHSCVDLDDKGCFAEQRASPLTSISAVVG 660
Db 601 PSQVKPDLSPYMPWKFPDEBEGACQPCINCTHSCVDLDDKGCFAEQRASPLTSISAVVG 660
Qy 661 ILLVVLGVVFGILIKRQOKIRKYTMRLLOLTELVEPLTPSGAMPNQAMRILKETEL 720
Db 661 ILLVVLGVVFGILIKRQOKIRKYTMRLLOLTELVEPLTPSGAMPNQAMRILKETEL 720
Qy 721 RKVKVLGSGAFGVYKGIWIPDGENVKI PVAIKVLRNTSPKANKETLDEAYVMAGVSP 780
Db 721 RKVKVLGSGAFGVYKGIWIPDGENVKI PVAIKVLRNTSPKANKETLDEAYVMAGVSP 780

Db	1141	NQDVRQPPSPREGPLPAARPAAGATLERKTLSPGKNGVVKDVFAGGAVERPEYLTTPQ	1200
QY	1201	GGAAPQHPPPAFSPAFDNLYYWDQDPPRGAPSTFKGTPTAENPEYLGIDVPV	1255
Db	1201	GGAAPQHPPPAFSPAFDNLYYWDQDPPRGAPSTFKGTPTAENPEYLGIDVPV	1255
RESULT 12			
ID	AAE20479	standard; Protein; 1255 AA.	
XX	AAE20479;		
AC	01-JUL-2002	(first entry)	
DT			
XX		Human Her-2/neu protein.	
XX			
XX		Human; Her-2/Neu protein; immune response; gene therapy; breast cancer;	
KW		human leukocyte antigen; HLA; vaccine; malignancy; cytostatic.	
XX			
OS		Homo sapiens.	
XX			
FH	Key	Location/Qualifiers	
FT	Region	1021..1030	
FT		/note="Naturally processed HLA-B44-restricted epitope"	
XX	WO200214503-A2.		
PN	21-FEB-2002.		
XX			
XX	14-AUG-2001;	2001WO-US41733.	
PF			
XX	14-AUG-2000;	2000US-225152P.	
PR	28-SEP-2000;	2000US-236428P.	
PR	21-FEB-2001;	2001US-270520P.	
XX			
XX	(CORI-) CORIXA CORP.		
PA			
PI	Hand-zimmermann S, Cheever MA, Foy TM, Lodes MJ, Kalos MD;		
PI	Mcneill PD, Vedvick TS;		
XX			
DR	WPI; 2002-280758/32.		
DR	N-PSDB; AAD32743.		
XX			
PT	Novel isolated Her-2/Neu polypeptide composition useful for therapy,		
PT	prevention and diagnosis of cancer, preferably breast cancer		
PS	Disclosure; Page 114-117; 129pp; English.		
XX			
CC	The invention relates to an isolated Her-2/Neu polypeptide composition effective for eliciting an immune response. The invention is useful for eliciting an immune response in a patient, where the patient is human leukocyte antigen (HLA)-B44 positive or is affected with breast cancer. The composition is useful for the therapy and diagnosis of cancer, preferably breast cancer, in pharmaceutical compositions, e.g., vaccine and other compositions for the diagnosis, prevention and treatment of human malignancies, for stimulating and/or expanding T cells specific for Her-2/Neu polypeptide and for inhibiting the development of cancer in a patient. The invention is useful for stimulating a T cell response in a human patient, as probe or primer for nucleic acid hybridisation, to selectively form duplex molecules with complementary stretches of the entire Her-2/Neu gene or gene fragments of interest, to isolate a full length gene from a suitable library, and to direct expression of a polypeptide in appropriate host cells. The composition is useful in prophylactic or therapeutic applications and for the treatment of cancer, preferably for the immunotherapy of breast cancer and other Her-2/Neu-associated malignancies. The invention is useful in gene therapy. The present sequence is human Her-2/neu protein.		
XX	Sequence	1255 AA;	

Query Match 99.9%; Score 6806; DB 23; Length 1255;

Best Local Similarity 99.8%; Pred. No. 0;			
Matches 1253; Conservative 1; Mismatches 1; Indels 0; Gaps 0;			
QY	1	MELAAALCRWGLLALALPPGAASQVCTGDMKMLRLPASPTHLDMLRHLHYQGQVVOGNL	60
Db	1	MELAAALCRWGLLALALPPGAASQVCTGDMKMLRLPASPTHLDMLRHLHYQGQVVOGNL	60
QY	61	ELTYLFTNASLSFLQDIEQVQGVYLIHAHQVQVPLQRLRIVRGTQLFEDNYALAVLDNG	120
Db	61	ELTYLFTNASLSFLQDIEQVQGVYLIHAHQVQVPLQRLRIVRGTQLFEDNYALAVLDNG	120
QY	121	DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDTILWKDIFHKQNOLA	180
Db	121	DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDTILWKDIFHKQNOLA	180
QY	181	LTLIDTNRSRACHPCSPMCKGSCWGESSEDCQSLRTVCAGGCARCKGPLPTDCCHEQC	240
Db	181	LTLIDTNRSRACHPCSPMCKGSCWGESSEDCQSLRTVCAGGCARCKGPLPTDCCHEQC	240
QY	241	AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMENPEGRYTFGASCVTACP	300
Db	241	AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMENPEGRYTFGASCVTACP	300
QY	301	YNYLSTDVGSCTLVCPHNOEVTAEADGTORCEKSKPCARVCYGLGMEHLREVRVTSAN	360
Db	301	YNYLSTDVGSCTLVCPHNOEVTAEADGTORCEKSKPCARVCYGLGMEHLREVRVTSAN	360
QY	361	IOEFAGCKKIFGSLAPLPESFDGDPASNTAPLOPELOVQFETLEEITGVLYISAWPDSL	420
Db	361	IOEFAGCKKIFGSLAPLPESFDGDPASNTAPLOPELOVQFETLEEITGVLYISAWPDSL	420
QY	421	DLVSFQNLQVIRGRILHNGAYSLTLQGLGSIWGLRSLRSLRELASGLALIHNTLFCFVHTV	480
Db	421	DLVSFQNLQVIRGRILHNGAYSLTLQGLGSIWGLRSLRSLRELASGLALIHNTLFCFVHTV	480
QY	481	PWDQLFRNPHQALLHTANRPEDECVGEGACHQLCARGHCWGPGPTQCVNCSQFLRGQEC	540
Db	481	PWDQLFRNPHQALLHTANRPEDECVGEGACHQLCARGHCWGPGPTQCVNCSQFLRGQEC	540
QY	541	VVECRVLQGLPREYVNAHCLPCHPECPONGSVTCFGEADOCVACAHYKDPFCVCAR	600
Db	541	VVECRVLQGLPREYVNAHCLPCHPECPONGSVTCFGEADOCVACAHYKDPFCVCAR	600
QY	601	PSGVKPDLSYMPIMKFPDEEGACQPCPINCTHSCVDLDDKGPAPQASPLTSIVSAVVG	660
Db	601	PSGVKPDLSYMPIMKFPDEEGACQPCPINCTHSCVDLDDKGPAPQASPLTSIVSAVVG	660
QY	661	ILLVVVLGVVFGILIKRQOKIRKYTMRLLOETELVEPLTPSGAMPNQAQMRILKETEL	720
Db	661	ILLVVVLGVVFGILIKRQOKIRKYTMRLLOETELVEPLTPSGAMPNQAQMRILKETEL	720
QY	721	RKVKVLGSGAFGVYKGIWIPGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGVGP	780
Db	721	RKVKVLGSGAFGVYKGIWIPGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGVGP	780
QY	781	YVSRLLGICLTSTVOLVTOLMPYGLLDHVRNRRGLSQDILLNWCMLAKGMSYLEDVR	840
Db	781	YVSRLLGICLTSTVOLVTOLMPYGLLDHVRNRRGLSQDILLNWCMLAKGMSYLEDVR	840
QY	841	LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPFKMMALESILRRRFT	900
Db	841	LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPFKMMALESILRRRFT	900
QY	901	HQSDVMSYGVTVWELMTFGAKYDGI PAREIPDLLEKGERLPQPPCTTIDVYMWKCM	960
Db	901	HQSDVMSYGVTVWELMTFGAKYDGI PAREIPDLLEKGERLPQPPCTTIDVYMWKCM	960
QY	961	IDSECRPRFRELVSFSEFMRARDPQRFVVTQNEIDLGPASPLDSTFFVRSLEDDDMGLVDA	1020
Db	961	IDSECRPRFRELVSFSEFMRARDPQRFVVTQNEIDLGPASPLDSTFFVRSLEDDDMGLVDA	1020
QY	1021	EEYLVPPQGGFFCPDPAPGAGGMVHRHRSSTRSGGDLTLGLEPSEEPAPSLPASEG	1080

Db 1141 NQDVRQPSRSGPLPAARPAAGATLERPKTSLSPKNGVVKDVPARFGGAVENPEYLTPQ 1200
Qy 1201 GGAAPQPHPPAFSPAFDNLYWQDPPPERGAPSTFKGTPTAENPEYGLDVPV 1255
Db 1201 GGAAPQPHPPAFSPAFDNLYWQDPPPERGAPSTFKGTPTAENPEYGLDVPV 1255

RESULT 10

AAG88267
ID AAG88267 standard; Protein; 1255 AA.

XX AC AAG88267;

XX DT 11-SEP-2001 (first entry)

XX DE HER2/neu amino acid sequence.

XX KW Human; HER2/neu; epitope; human leukocyte antigen; HLA; T cell;
XX KW immune response; vaccine; tumour; cancer; cytotoxic; immunostimulant;
XX KW tumour-associated antigen; T lymphocyte; cytotoxic T lymphocyte; CTL.

XX OS Homo sapiens.

XX PN WO200141787-A1.

XX PD 14-JUN-2001.

XX PF 11-DEC-2000; 2000WO-US33591.

XX PR 10-DEC-1999; 99US-0458299.

XX PA (EPIM-) EPIMUNE INC.

XX PI Fikes J, Sette A, Sidney J, Southwood S, Chesnut R, Celis E;

XX PI Keogh E;

XX DR WPI; 2001-374995/39.

XX PT An isolated prepared HER2/neu epitope useful in a vaccine for inducing
XX PT cellular immune responses for the prevention and treatment of cancer -
XX PS Disclosure; Page 15; 199pp; English.

XX CC The present invention describes isolated prepared HER2/neu epitopes (I).
XX CC Also described are: (1) a clonal cytotoxic T lymphocyte (CTL) that is
XX CC culture in vitro and binds to a complex of an epitope (I), bound to a
XX CC human leukocyte antigen (HLA) molecule; (2) a peptide (II) comprising (I)
XX CC and a second epitope and the peptide is less than 50 contiguous amino
XX CC acids that have 100% identity with a native peptide sequence of HER2/neu;
XX CC (3) a vaccine composition (III) comprising (II) and a pharmaceutical
XX CC excipient; (4) an isolated nucleic acid encoding a peptide comprising
XX CC (I); and (5) an isolated nucleic acid encoding (II). (I) has cytostatic
XX CC and immunostimulant activities, and can be used in vaccines. (I), (II)
XX CC and (III) are useful for inducing cellular immune responses for the
XX CC prevention and treatment of cancer. (I) and (II) are useful for
XX CC monitoring or evaluating an immune response to a tumour-associated
XX CC antigen when incubated with a T lymphocyte sample from a patient and
XX CC detecting the presence of bound T lymphocyte to (I) or (II). Epitope
XX CC based vaccines mean that immunosuppressive epitopes that may be present
XX CC in whole antigens may be avoided. Selected epitopes may be combined to
XX CC enhance immunogenicity. The possible pathological side effects caused by
XX CC infectious agents or whole protein antigen is eliminated. The vaccine
XX CC provides the ability to direct and focus an immune response to multiple
XX CC selected antigens from the same pathogen. Epitope-based anti-tumour
XX CC vaccines provides the opportunity to combine epitopes derived from
XX CC multiple tumour-associated molecules addressing the problem of tumour-
XX CC tumour variability and reducing the likelihood of tumour escape due to
XX CC antigen loss. AAG88266 to AAG89121 represent amino acid sequences used in
XX CC the exemplification of the present invention.

XX SQ Sequence 1255 AA;

Query Match 99.7%; Score 6698; DB 22; Length 1255;

Best Local Similarity 98.8%; Pred. No. 0;
Matches 1240; Conservative 1; Mismatches 14; Indels 0; Gaps 0;
Qy 1 MELAALCRWGLLALLPPGAASTQVCTGTDMLRLPASPTHLDMLRLHLYQGQVVOGNL 60
Db 1 MELAALCRWGLLALLPPGAASTQVCTGTDMLRLPASPTHLDMLRLHLYQGQVVOGNL 60
Qy 61 ELTYLPTNASLSFLQDIQEVQGYVLIJAHNOVRQVPLQRLRIVRGTOQLFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIQEVQGYVLIJAHNOVRQVPLQRLRIVRGTOQLFEDNYALAVLDNG 120
Qy 121 DPLNNTTPTVGASPGGLRELQRLSLTEILKGGVLIQBNPOLCYQDTILWKDIFHKNNOLA 180
Db 121 DPLNNTTPTVGASPGGLRELQRLSLTEILKGGVLIQBNPOLCYQDTILWKDIFHKNNOLA 180
Qy 181 LTLIDTNRSRACHPCSPMCKSGRCWGESSEDCOSLTRTVCAGGCARCKGKPLPOVIKANSK 240
Db 181 LTLIDTNRSRACHPCSPMCKSGRCWGESSEDCOSLTRTVCAGGCARCKGKPLPDDCCEHCQ 240
Qy 241 FIGITELKSHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMPNPBGRTYTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMPNPBGRTYTFGASCVTACP 300
Qy 301 YNYLSTDVGSCTLVCPHLNQEVTAEDTORCEKSKPCARVCYGLGMEHLREVRVAVTSAN 360
Db 301 YNYLSTDVGSCTLVCPHLNQEVTAEDTORCEKSKPCARVCYGLGMEHLREVRVAVTSAN 360
Qy 361 IOEFAGCKIFGSLAFPLPSFDGDPASNTAPLOEQVPELTLEETIGYLIISAWPDSLP 420
Db 361 IOEFAGCKIFGSLAFPLPSFDGDPASNTAPLOEQVPELTLEETIGYLIISAWPDSLP 420
Qy 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGISWGLSLRELGLSLALIHNNTHLCFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGISWGLSLRELGLSLALIHNNTHLCFVHTV 480
Qy 481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHOLCARGHCWGPGTQCVNCSFRLRQEC 540
Db 481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHOLCARGHCWGPGTQCVNCSFRLRQEC 540
Qy 541 VECRVLQGLPREYVVARHCLPCHPCOPONGSVTCFGEADOCVCAHYKDPFPCVARC 600
Db 541 VECRVLQGLPREYVVARHCLPCHPCOPONGSVTCFGEADOCVCAHYKDPFPCVARC 600
Qy 601 PSGVKPDLSTYMPKPFDEEGACQPCPINCTHSCVDLDDKGCFAEQRASPLTSISAVVG 660
Db 601 PSGVKPDLSTYMPKPFDEEGACQPCPINCTHSCVDLDDKGCFAEQRASPLTSISAVVG 660
Qy 661 ILLVVLGVVFGILIKRROOKIRKYTMRLLOQTELVLEPLTPSGAMPNQMRILKETEL 720
Db 661 ILLVVLGVVFGILIKRROOKIRKYTMRLLOQTELVLEPLTPSGAMPNQMRILKETEL 720
Qy 721 RKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKETLDEAYVMAGVGP 780
Db 721 RKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKETLDEAYVMAGVGP 780
Qy 781 YVSRLLGICLTSTVQLVTQMLPMPYGLLDHVNRNRLGSGQDLNLCWQIAKMSYLEDVR 840
Db 781 YVSRLLGICLTSTVQLVTQMLPMPYGLLDHVNRNRLGSGQDLNLCWQIAKMSYLEDVR 840
Qy 841 LVHRDLAARNVLKSPNNHVKITDFTGLARLLDDETHYADGGKVPKKNMALESILRRRFT 900
Db 841 LVHRDLAARNVLKSPNNHVKITDFTGLARLLDDETHYADGGKVPKKNMALESILRRRFT 900
Qy 901 HQSDVMSYGVYVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYMIWKVCM 960
Db 901 HQSDVMSYGVYVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYMIWKVCM 960
Qy 961 IDSECEPRFRELVSERFARMARDPQRFVJQNEDLGPASPLDSTFYRSLLEDDDMGLVDA 1020
Db 961 IDSECEPRFRELVSERFARMARDPQRFVJQNEDLGPASPLDSTFYRSLLEDDDMGLVDA 1020
Qy 1021 BEYLVFQQGFCCPDPAAGAGMWHHRSSSTRSGGDLTLGLPESEBAPRPLAPSEG 1080

Db 1021 EYLVPQQGFFCPDPAPGAGWVHRRSSSTRSGGDLTLGLEPSEEAAPRSLAPSEG 1080
 Qy 1081 AGSDVFDGLGMAKGLQSLPTDPSLQRYSEDPTVLPSETDGVVAPLTCSPQPEYV 1140
 Db 1081 AGSDVFDGLGMAKGLQSLPTDPSLQRYSEDPTVLPSETDGVVAPLTCSPQPEYV 1140
 Qy 1141 NQDVRPQPPSPREGPLPAARPAAGATLERAKTLLSPGKNGVVKDVFAFGGAVENPEYLTPO 1200
 Db 1141 NQDVRPQPPSPREGPLPAARPAAGATLERPKTLLSPGKNGVVKDVFAFGGAVENPEYLTPO 1200
 Qy 1201 GGAAPQHPHPPAFSPADNLYWQDPPPERGAPSTFKGTPTAENPYLGLDVPV 1255
 Db 1201 GGAAPQHPHPPAFSPADNLYWQDPPPERGAPSTFKGTPTAENPYLGLDVPV 1255

RESULT 11

AAE24067
 ID . AAE24067 standard; Protein; 1255 AA.

AC AAE24067;

DT 23-SEP-2002 (first entry)

DE Human Her-2 protein.

XX Human; Her-2; epidermal growth factor receptor 2; infection; cancer;
 KW hyperproliferative disorder; prophylaxis; inflammation; antisense;
 KW tumour; gene therapy; phosphorothioate backbone.

XX Homo sapiens.

XX WO200222636-A1.

XX 21-MAR-2002.

XX 12-SEP-2001; 2001WO-US28572.

XX 15-SEP-2000; 2000US-0663834.

XX (ISIS-) ISIS PHARM INC.

XX Bennett CF, Cowser LM;

XX WPI; 2002-471192/50.

XX N-PSDB; RAD38904.

XX Novel antisense oligonucleotide which modulates the expression of Human
 PT Epidermal Growth Factor receptor, Her2, is useful for treating tumors
 PT inflammation or to prevent infection in humans -

XX Example 13; Page 95-107; 116pp; English.

XX The invention relates to antisense compounds targeted to a nucleic
 CC acid molecule encoding Her2 (human Epidermal Growth Factor receptor 2)
 CC that specifically hybridizes with and inhibits the expression of Her2.
 CC Antisense compounds of the invention are used for treating diseases or
 CC conditions associated with Her2 such as hyperproliferative disorders
 CC e.g. lung, breast, gastric, oesophageal, colon, bladder, salivary,
 CC neural or cardiac cancer. They are also useful prophylactically e.g.
 CC to prevent or delay infection, inflammation and tumour formation. The
 CC invention is also used in gene therapy. The present sequence is human
 CC Her-2 protein.

XX Sequence 1255 AA;

Query Match 98.7%; Score 6698; DB 23; Length 1255;

Best Local Similarity 98.8%; Pred. No. 0;

Matches 1240; Conservative 1; Mismatches 14; Indels 0; Gaps 0;

Qy 1 MELAALCRWGLLLALLPPGAASCTGCTDMKILPLASPETHLDMRLHYGCGVQGNL 60

Db 1 MELAALCRWGLLLALLPPGAASCTGCTDMKILPLASPETHLDMRLHYGCGVQGNL 60

Qy 61 ELTYLPTNASLSPLODIOEQVQYVLIHAHQVQVPLQRLRIVRGTQOLFEDNYALAVLDNG 120
 Db 61 ELTYLPTNASLSPLODIOEQVQYVLIHAHQVQVPLQRLRIVRGTQOLFEDNYALAVLDNG 120
 Qy 121 DPLNNTPTVTGASPGGLRELQRLASLTELKGGVLIQORNPOLCYQDITLWKDIFHKKNOLA 180
 Db 121 DPLNNTPTVTGASPGGLRELQRLASLTELKGGVLIQORNPOLCYQDITLWKDIFHKKNOLA 180
 Qy 181 LTLIDTNRACHPCSPCKSGRCWGESSEDCOSLRTVCAGGCARCKGPLPOYIKANSK 240
 Db 181 LTLIDTNRACHPCSPCKSGRCWGESSEDCOSLRTVCAGGCARCKGPLPDDCCHCEQC 240
 Qy 241 FIGITELKISDCLACLFHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
 Db 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
 Qy 301 YNYLSTDVGSCTLVCPHNOEVTAEQTCRCKSPCARVCYGLGMEHLREVRVTSAN 360
 Db 301 YNYLSTDVGSCTLVCPHNOEVTAEQTCRCKSPCARVCYGLGMEHLREVRVTSAN 360
 Qy 361 IOEFAGCKIFGSLAFLPESFDGDPASNTAPLOEQVFEETLEETGYLIISAWPDSL 420
 Db 361 IOEFAGCKIFGSLAFLPESFDGDPASNTAPLOEQVFEETLEETGYLIISAWPDSL 420
 Qy 421 DLSVFQNLQVIRGRIILHNGAYSITLQGLGTSWGLSLRELGLSLALIHNTHLFCFVHTV 480
 Db 421 DLSVFQNLQVIRGRIILHNGAYSITLQGLGTSWGLSLRELGLSLALIHNTHLFCFVHTV 480
 Qy 481 PWDOLFNRHQALLHTANRPEDECVGEGLAACHOLCARGHCWGPGPTQVCNCSOFLRGQEC 540
 Db 481 PWDOLFNRHQALLHTANRPEDECVGEGLAACHOLCARGHCWGPGPTQVCNCSOFLRGQEC 540
 Qy 541 VEECRVLQGLPREYVNRHCLPCHPECOQNGSVTCFGEADOCVCAHYKDPFCVARC 600
 Db 541 VEECRVLQGLPREYVNRHCLPCHPECOQNGSVTCFGEADOCVCAHYKDPFCVARC 600
 Qy 601 PSGVKPDLSPYMPKFPDEEGACQPCPINCTHSCVDLDDKGCAPAEQASPLTSISAVVG 660
 Db 601 PSGVKPDLSPYMPKFPDEEGACQPCPINCTHSCVDLDDKGCAPAEQASPLTSISAVVG 660
 Qy 661 ILLVVLGVVFGILIKRQOKIRKTYMRRLLQSTELVEPLTPSGAMPNQAMRILKETEL 720
 Db 661 ILLVVLGVVFGILIKRQOKIRKTYMRRLLQSTELVEPLTPSGAMPNQAMRILKETEL 720
 Qy 721 RKVKVLGSGAGFVYKGIWIPDGENVKIPVAIKVLRNTSPKANKETLDEAYVMAGVSP 780
 Db 721 RKVKVLGSGAGFVYKGIWIPDGENVKIPVAIKVLRNTSPKANKETLDEAYVMAGVSP 780
 Qy 781 YVSRLLGICLTSTVQLVTQMLPYGCLLDHVRENRGLSGQDLLNWCQIAKGNYSILEDVR 840
 Db 781 YVSRLLGICLTSTVQLVTQMLPYGCLLDHVRENRGLSGQDLLNWCQIAKGNYSILEDVR 840
 Qy 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESILRRRFT 900
 Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESILRRRFT 900
 Qy 901 HQSDVWSYGVYVWELMTFGAKPYDGIIPAREIPLLEKGERLPQPPCTIDVYIMVWKCM 960
 Db 901 HQSDVWSYGVYVWELMTFGAKPYDGIIPAREIPLLEKGERLPQPPCTIDVYIMVWKCM 960
 Qy 961 IDSECRPRFRELVSERFARMARDPQRFVVIQNEIDLGASPLDSTFYRSLLEDDMDGLVDA 1020
 Db 961 IDSECRPRFRELVSERFARMARDPQRFVVIQNEIDLGASPLDSTFYRSLLEDDMDGLVDA 1020
 Qy 1021 EYLVFPQQGFFCPDPAPGAGWVHRRSSSTRSGGDLTLGLEPSEEAAPRSLAPSEG 1080
 Db 1021 EYLVFPQQGFFCPDPAPGAGWVHRRSSSTRSGGDLTLGLEPSEEAAPRSLAPSEG 1080
 Qy 1081 AGSDVFDGLGMAKGLQSLPTDPSLQRYSEDPTVLPSETDGVVAPLTCSPQPEYV 1140
 Db 1081 AGSDVFDGLGMAKGLQSLPTDPSLQRYSEDPTVLPSETDGVVAPLTCSPQPEYV 1140
 Qy 1141 NQDVRPQPPSPREGPLPAARPAAGATLERAKTLLSPGKNGVVKDVFAFGGAVENPEYLTPO 1200

Db 1141 NQPDVRQPPSPREGPLPAARPAATLERPKTLSPGKGVVVDVFAFGGAVENPEYLTPO 1200
QY 1201 GGAAPQHPPPAFSPADNLYYWDQPPPERGAPSTFKGTPTAENPEYLGIDVPV 1255
Db 1201 GGAAPQHPPPAFSPADNLYYWDQPPPERGAPSTFKGTPTAENPEYLGIDVPV 1255
RESULT 12
AAE20479
ID AAE20479 standard; Protein; 1255 AA.
XX
AC AAE20479;
DT 01-JUL-2002 (first entry)
XX
DE Human Her-2/neu protein.
XX
KW Human; Her-2/Neu protein; immune response; gene therapy; breast cancer;
KW human leukocyte antigen; HLA; vaccine; malignancy; cytostatic.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Region 1021..1030
FT /note= "Naturally processed HLA-B44-restricted epitope"
XX
PN WO200214503-A2.
XX
PD 21-FEB-2002.
XX
PF 14-AUG-2001; 2001WO-US41733.
XX
PR 14-AUG-2000; 2000US-225152P.
PR 28-SEP-2000; 2000US-236428P.
PR 21-FEB-2001; 2001US-270520P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Hand-zimmermann S, Cheever MA, Foy TM, Lodes MJ, Kalos MD;
PI McNeill PD, Vedvick TS;
XX
XX WFI; 2002-280758/32.
DR N-PSDB; AAD32743.
XX
XX Novel isolated Her-2/Neu polypeptide composition useful for therapy,
PT prevention and diagnosis of cancer, preferably breast cancer -
PT
XX Disclosure; Page 114-117; 129pp; English.
XX
XX The invention relates to an isolated Her-2/Neu polypeptide composition
CC effective for eliciting an immune response. The invention is useful for
CC eliciting an immune response in a patient, where the patient is human
CC leukocyte antigen (HLA)-B44 positive or is affected with breast cancer.
CC The composition is useful for the therapy and diagnosis of cancer,
CC preferably breast cancer, in pharmaceutical compositions, e.g., vaccine
CC and other compositions for the diagnosis, prevention and treatment of
CC human malignancies, for stimulating and/or expanding T cells specific for
CC Her-2/Neu polypeptide and for inhibiting the development of cancer in a
CC patient. The invention is useful for stimulating a T cell response in a
CC human patient, as probe or primer for nucleic acid hybridisation, to
CC selectively form duplex molecules with complementary stretches of the
CC entire Her-2/Neu gene or gene fragments of interest, to isolate a full
CC length gene from a suitable library, and to direct expression of a
CC polypeptide in appropriate host cells. The composition is useful in
CC prophylactic or therapeutic applications and for the treatment of cancer,
CC preferably for the immunotherapy of breast cancer and other Her-2/Neu-
CC associated malignancies. The invention is useful in gene therapy. The
CC present sequence is human Her-2/neu protein.
XX
SQ Sequence 1255 AA;

Query Match

98.7%; Score 6698; DB 23; Length 1255;

Best Local Similarity 98.8%; Pred. No. 0;
Matches 1240; Conservative 1; Mismatches 14; Indels 0; Gaps 0;
QY 1 MELAALCRWGLLALLPPGAASQVCTGTDMLRLPASPTHLDMLRHLYQGQVVOGNL 60
Db 1 MELAALCRWGLLALLPPGAASQVCTGTDMLRLPASPTHLDMLRHLYQGQVVOGNL 60
QY 61 ELYLPTNASLSFLQDIQEVQGYVLIHAHQVQVPLQRLIRIVRGTOQLFEDNYALAVLDNG 120
Db 61 ELYLPTNASLSFLQDIQEVQGYVLIHAHQVQVPLQRLIRIVRGTOQLFEDNYALAVLDNG 120
QY 121 DPLNNTTPTVGASPGGIRELOLRSLTEILKGGVLIQORNPOLCYODILKWKDI FHKNNOLA 180
Db 121 DPLNNTTPTVGASPGGIRELOLRSLTEILKGGVLIQORNPOLCYODILKWKDI FHKNNOLA 180
QY 181 LTLIDTNRGRACHPCSPCKSGCWGESSEDCOSLTRITVCAGGCARCKGPLPOYIKANSK 240
Db 181 LTLIDTNRGRACHPCSPCKSGCWGESSEDCOSLTRITVCAGGCARCKGPLPDDCCEHC 240
QY 241 FIGITELKHSDDLACILHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Db 241 AAGCTGPKHSDDLACILHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
QY 301 YNYLSTDVSGSCTLVCPHLHQEVTAEDGTORCEKSKPCARVCYGLGMEHLREVRVTSAN 360
Db 301 YNYLSTDVSGSCTLVCPHLHQEVTAEDGTORCEKSKPCARVCYGLGMEHLREVRVTSAN 360
QY 361 IQEFAGCKIFGSLAFPLPESFDGDPASNTAPLOQVPELTLEETITGYLYISAWPDSLP 420
Db 361 IQEFAGCKIFGSLAFPLPESFDGDPASNTAPLOQVPELTLEETITGYLYISAWPDSLP 420
QY 421 DLSVFQNLQVIRGRILHNGAYSITLQGLGISMGLSLRELGSGLALIHNTHLFCFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYSITLQGLGISMGLSLRELGSGLALIHNTHLFCFVHTV 480
QY 481 PWDQLFRNPHQALLHTANRPEDCEVGEGLACHOLCARGHCWGPGPTQCVNCSOFLRQEC 540
Db 481 PWDQLFRNPHQALLHTANRPEDCEVGEGLACHOLCARGHCWGPGPTQCVNCSOFLRQEC 540
QY 541 VEECRVLOGLPREYVNRHCLPCHPECOQNGSVTCFGEADOCVCAHYKDPFCVARC 600
Db 541 VEECRVLOGLPREYVNRHCLPCHPECOQNGSVTCFGEADOCVCAHYKDPFCVARC 600
QY 601 PSGVKPDLSPYMPIWKFPEDEGACQPCINCTHSCVDLDDKGCPEARASPLTSIVSAVVG 660
Db 601 PSGVKPDLSPYMPIWKFPEDEGACQPCINCTHSCVDLDDKGCPEARASPLTSIVSAVVG 660
QY 661 ILLVVVLGVVFGILIKRROQKIRKYTMRLLOSTELVEPLTPSGAMPNOAQMRLKETEL 720
Db 661 ILLVVVLGVVFGILIKRROQKIRKYTMRLLOSTELVEPLTPSGAMPNOAQMRLKETEL 720
QY 721 RKVKVLGSGAGFVYVKGWIPOGENVKIPVAIKVLRNTSPKANKIIDEAYVMAGVGP 780
Db 721 RKVKVLGSGAGFVYVKGWIPOGENVKIPVAIKVLRNTSPKANKIIDEAYVMAGVGP 780
QY 781 YVSRLLGICLTSTVQLVTOLMPYGLLDHVNRNRLGSDQLLNWCQIAKGMYSILEDVR 840
Db 781 YVSRLLGICLTSTVQLVTOLMPYGLLDHVNRNRLGSDQLLNWCQIAKGMYSILEDVR 840
QY 841 LVHRDLAARNVLKSPNNHVKITDFGLARLLDIDETEHADGGKVPKWKMALESILRRRFT 900
Db 841 LVHRDLAARNVLKSPNNHVKITDFGLARLLDIDETEHADGGKVPKWKMALESILRRRFT 900
QY 901 HQSDVMSYGVYVWELMTFGAKPYDGI PAREI PDLEKGERLPPOPPICTIDVYIMVKCWM 960
Db 901 HQSDVMSYGVYVWELMTFGAKPYDGI PAREI PDLEKGERLPPOPPICTIDVYIMVKCWM 960
QY 961 IDSECRPRFRELVSFSESRMARDPQRFWITQNEDLGPASPLDSTFYRSLDEDDMGDLVDA 1020
Db 961 IDSECRPRFRELVSFSESRMARDPQRFWITQNEDLGPASPLDSTFYRSLDEDDMGDLVDA 1020
QY 1021 BEYLVPOQGFPCDPAPGAGGMVHHRRSSSTRSGGDLTLGLPESEEBAPSLAPSEG 1080
Db 1021 BEYLVPOQGFPCDPAPGAGGMVHHRRSSSTRSGGDLTLGLPESEEBAPSLAPSEG 1080

Db 1021 EEVLVPOQGFCCPDPAPGAGWVHRRSSSTRSGGDLTLGLEPSEEEAPRSPAPSEG 1080
 Qy 1081 AGSDVFDGDLGMAKGLQSLPHDPSPLQRYSEDPTVPLPSETDGVVAPITCSPQPEYV 1140
 Db 1081 AGSDVFDGDLGMAKGLQSLPHDPSPLQRYSEDPTVPLPSETDGVVAPITCSPQPEYV 1140
 Qy 1141 NQPDVRQPSPREGPLPAARPAAGATLERAKTLSPGKNGVGVKOVFAFGGAVENPEYLTPO 1200
 Db 1141 NQPDVRQPSPREGPLPAARPAAGATLERAKTLSPGKNGVGVKOVFAFGGAVENPEYLTPO 1200
 Qy 1201 GGAAPQHPPPAFSPADNLYYWDQDPPERGAPPSTFKGTPTAENPYLGLDVPV 1255
 Db 1201 GGAAPQHPPPAFSPADNLYYWDQDPPERGAPPSTFKGTPTAENPYLGLDVPV 1255

RESULT 13

AAMS1143
 ID AAMS1143 standard; Protein; 1255 AA.

XX AC AAMS1143;
 XX 17-JUN-2002 (first entry)
 XX Human Her-2/neu oncogene-encoded p185 glycoprotein.
 XX Her-2/neu; oncogene; cancer; tumour; vaccine; human; p185;
 XX tyrosine kinase; receptor; c-erbB2; gene therapy.
 XX Homo sapiens.
 XX Key Location/Qualifiers
 XX Domain 1..653
 XX /note= "extracellular domain"
 XX Domain 676..1255
 XX /note= "intracellular domain"
 XX Domain 990..1255
 XX /note= "phosphorylation domain"
 XX W0200212341-A2.
 XX 14-FEB-2002.
 XX 03-AUG-2001; 2001WO-US24283.
 XX 03-AUG-2000; 2000US-0632507.
 XX (CORI-) CORIXA CORP.
 XX (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
 XX Cheever MA, Gheysen D;
 XX WPI; 2002-241743/29.
 XX N-PSDB; ABA92250.

Her-2/neu fusion protein for treating or preventing cancer by eliciting
 or enhancing an immune response to the protein, has Her-2/neu
 extracellular domain fused to Her-2/neu intracellular or
 phosphorylation domain
 Claim 68; Fig 7; 141pp; English.

The present sequence is that of human Her-2/neu (p185 glycoprotein
 or c-erbB2), an oncogenic self-protein and target for anti-cancer
 vaccines. The Her-2/neu gene is amplified and p185 is overexpressed
 in a variety of cancers, including breast, ovarian, colon, lung and
 prostate cancer. Her-2/neu is a member of the tyrosine kinase
 family of receptor-like glycoproteins. It comprises an extracellular
 domain with homology to the epidermal growth factor receptor
 (EGFR), a highly hydrophobic transmembrane domain and a C-terminal
 intracellular domain that also shows homology to EGFR. Its
 overexpression correlates with a poor prognosis in breast and
 ovarian cancers. The invention provides Her-2/neu fusion
 proteins, nucleic acids encoding them, viral vectors, and vaccines

comprising the fusion proteins or nucleic acid molecules. In
 preferred fusion proteins, the extracellular domain of a Her-2/neu
 protein is fused to a Her-2/neu intracellular domain or
 phosphorylation domain (or its DeltaPD fragment). An immune
 response to Her-2/neu protein is elicited or enhanced by
 administering the fusion protein in the form of a vaccine, or by
 transfecting cells of an animal ex vivo with a nucleic acid
 encoding the fusion protein, and delivering the transfected cells
 to the animal. The fusion proteins, nucleic acids, and isolated
 specific T-cells are useful for inhibiting the development of a
 cancer, especially breast, ovarian, colon, lung or prostate cancer
 in a patient. T cells that specifically react with a Her-2/neu
 fusion protein can be used to remove tumour cells from a sample in
 order to inhibit the development of cancer in a patient.

XX Sequence 1255 AA;

Query Match 98.7%; Score 6698; DB 23; Length 1255;
 Best Local Similarity 98.8%; Pred. No. 0;
 Matches 1240; Conservative 1; Mismatches 14; Indels 0; Gaps 0;

Qy 1 MELAALCRWGLLLALLPPGAASTQVCTGTDMLRLPASPTHLDMLRHLYQGCVVQGNL 60
 Db 1 MELAALCRWGLLLALLPPGAASTQVCTGTDMLRLPASPTHLDMLRHLYQGCVVQGNL 60
 Qy 61 ELTYLPTNASLSFLQDIQEVQGVVLIHNOVQVPLQRLRIVRGTLQFEDNYALAVLDNG 120
 Db 61 ELTYLPTNASLSFLQDIQEVQGVVLIHNOVQVPLQRLRIVRGTLQFEDNYALAVLDNG 120
 Qy 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDTILWKDIFHKNNQLA 180
 Db 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDTILWKDIFHKNNQLA 180
 Qy 181 LTLIDNRSRACHPCSPMCKSGRCWGESSEDCOSLRTVTCAGCAGCKGKPLQVIRANSK 240
 Db 181 LTLIDNRSRACHPCSPMCKSGRCWGESSEDCOSLRTVTCAGCAGCKGKPLQVIRANSK 240
 Qy 241 FIGITELKHSDDLACLHFNHSGICELHCPALVTYNTDTFESMPNPGRYTFGASCVTACP 300
 Db 241 AGCTGPKISDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPGRYTFGASCVTACP 300
 Qy 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSPCARVCYGLGMEHLREVRVNTSAN 360
 Db 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSPCARVCYGLGMEHLREVRVNTSAN 360
 Qy 361 IQEFAGCKIFGSLAFPLPESFDGDPASNTAPLOPELOVFETLEEITGYLIISAWPDSL 420
 Db 361 IQEFAGCKIFGSLAFPLPESFDGDPASNTAPLOPELOVFETLEEITGYLIISAWPDSL 420
 Qy 421 DLSVFQNLQVIRGRILHNGAYSLTLOGLGISWGLSLRLSRLGSLALIHNTLHLCFVHTV 480
 Db 421 DLSVFQNLQVIRGRILHNGAYSLTLOGLGISWGLSLRLSRLGSLALIHNTLHLCFVHTV 480
 Qy 481 PWDQLFRNPHQALLHTANRPDECVGEGGLACHOLCARGHCWGPGPTQCVNCSFLRQEC 540
 Db 481 PWDQLFRNPHQALLHTANRPDECVGEGGLACHOLCARGHCWGPGPTQCVNCSFLRQEC 540
 Qy 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFEGPEADQCVACAHKDPFPCVARC 600
 Db 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFEGPEADQCVACAHKDPFPCVARC 600
 Qy 601 PSGVKPDLSPVMPITWKPDEBACQPCPINTHSCVDLDDKGCAPAEORASPLTSTVSAVVG 660
 Db 601 PSGVKPDLSPVMPITWKPDEBACQPCPINTHSCVDLDDKGCAPAEORASPLTSTVSAVVG 660
 Qy 661 ILLVVLGVVFGILIKRQOKIRKYTMRLLQETELVEPLTPSGAMPNQAMRILKETEL 720
 Db 661 ILLVVLGVVFGILIKRQOKIRKYTMRLLQETELVEPLTPSGAMPNQAMRILKETEL 720
 Qy 721 RKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILD EAYVMAGVSP 780
 Db 721 RKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILD EAYVMAGVSP 780

QY 781 YVSRLLGICLTSTVQLVQLMPYGCCLLDHVRNRRGLSGDQLLNWCQIAKMSYLEDVR 840
DB 781 YVSRLLGICLTSTVQLVQLMPYGCCLLDHVRNRRGLSGDQLLNWCQIAKMSYLEDVR 840
QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMWMALESILRRFT 900
DB 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMWMALESILRRFT 900
QY 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYIMVKCWM 960
DB 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYIMVKCWM 960
QY 961 IDSECRPRFELVSESRMARDQRFVVIQNEIDLGPASPLDSTFYRSLLEDDMGDLVDA 1020
DB 961 IDSECRPRFELVSESRMARDQRFVVIQNEIDLGPASPLDSTFYRSLLEDDMGDLVDA 1020
QY 1021 EEVLPQOQFFCDDPAPAGAGWVHRRSSSTRSGGDLTLGLEPSEEEAPRPLAPSEG 1080
DB 1021 EEVLPQOQFFCDDPAPAGAGWVHRRSSSTRSGGDLTLGLEPSEEEAPRPLAPSEG 1080
QY 1081 AGSDVDFDGLGMAAGLQSLPTHDPSPLOQYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
DB 1081 AGSDVDFDGLGMAAGLQSLPTHDPSPLOQYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
QY 1141 NQDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAPGGAVENPEYLTQ 1200
DB 1141 NQDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAPGGAVENPEYLTQ 1200
QY 1201 GGAAPQHPPPAPSPAFDNLVYWDQDPPERGAPPSTFKGPTAENPEYLGDLVVP 1255
DB 1201 GGAAPQHPPPAPSPAFDNLVYWDQDPPERGAPPSTFKGPTAENPEYLGDLVVP 1255

RESULT 14

AAU77114
ID AAU77114 standard; Protein; 1255 AA.
XX AAU77114;
AC AAU77114;
XX
DT 05-JUN-2002 (first entry)
XX Human Her-2/neu polypeptide.
DE
XX Human; Her-2/neu; cytostatic; haematological malignancy; CML;
KW acute myelogenous leukaemia; AML; chronic myelogenous leukaemia; CLL;
KW chronic lymphocytic leukaemia; myeloma; non-Hodgkin's lymphoma; MDS;
KW Hodgkin's lymphoma; T cell therapy.
XX
OS Homo sapiens.
XX
XX WO200213847-A2.
XX
XX 21-FEB-2002.
XX
XX 13-AUG-2001; 2001WO-US25408.
XX
XX 14-AUG-2000; 2000US-0638280.
XX 28-SEP-2000; 2000US-0675904.
XX
XX (CORI-) CORIXA CORP.
XX
XX Gaiger A, Cheever MA, Hand-zimmermann S;
XX WPI; 2002-280741/32.
XX N-PSDB; ABK10730.
XX
XX Inhibiting haematological malignancy development by administering
PT polypeptide comprising immunogenic portion of Her-2/neu, polynucleotide
PT encoding the polypeptide, or antigen presenting cells expressing the
PT polypeptide
XX
XX Disclosure; Page 71-74; 74pp; English.
PS
XX

CC The invention relates to a method for inhibiting development of
CC haematological malignancy in a patient by administering a polypeptide
CC comprising an immunogenic portion of Her-2/neu or a polynucleotide
CC encoding the polypeptide. Antigen presenting cells that express the
CC protein can also be administered. The sequences are used for inhibiting
CC development of haematological malignancy such as acute myelogenous
CC leukaemia (AML), chronic myelogenous leukaemia (CML), chronic lymphocytic
CC leukaemia (CLL), MDS, myelomas, Hodgkin's lymphoma and non-Hodgkin's
CC lymphoma. This sequence represents the human Her-2/neu polypeptide.
XX
SQ Sequence 1255 AA;

Query Match 98.7%; Score 6698; DB 23; Length 1255;
Best Local Similarity 98.8%; Pred. No. 0;
Matches 1240; Conservative 1; Mismatches 14; Indels 0; Gaps 0;

QY 1 MELAALCRWGLLLALLPPGAASTQVCTGDMKRLRLPASPTHLDMLRHLYQGQVVOGNL 60
DB 1 MELAALCRWGLLLALLPPGAASTQVCTGDMKRLRLPASPTHLDMLRHLYQGQVVOGNL 60
QY 61 ELTYLPTNASLSFLQDIQEVQGVYLIHMQVRQVPLQRLIRVRGTQLFEDNYALAVLDNG 120
DB 61 ELTYLPTNASLSFLQDIQEVQGVYLIHMQVRQVPLQRLIRVRGTQLFEDNYALAVLDNG 120
QY 121 DPLNNTTPTVGASPGGLREQLRLSLTEILKGGVLIQORNPOLCYQDTILWKDIFHKNNOLA 180
DB 121 DPLNNTTPTVGASPGGLREQLRLSLTEILKGGVLIQORNPOLCYQDTILWKDIFHKNNOLA 180
QY 181 LTLIDTNRGRACHPCSPMCKGRCWGESSEDCOSLTRTVCAGGCARCKGPLPOYIKANSK 240
DB 181 LTLIDTNRGRACHPCSPMCKGRCWGESSEDCOSLTRTVCAGGCARCKGPLPTDCCHEQC 240
QY 241 FIGITELKHSDDLACLFHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
DB 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
QY 301 YNYLSTDVGSCTLVCPHMQEVTAEQGTORCKSKPCARVCYGLGMEHLREVRVAVTSAN 360
DB 301 YNYLSTDVGSCTLVCPHMQEVTAEQGTORCKSKPCARVCYGLGMEHLREVRVAVTSAN 360
QY 361 IOEFAGCKIKFSLAPLPESEFDGDPASNTAPLOPEQLQVFETLEETIGVLIYSAPDLSLP 420
DB 361 IOEFAGCKIKFSLAPLPESEFDGDPASNTAPLOPEQLQVFETLEETIGVLIYSAPDLSLP 420
QY 421 DLSVFQNLQVIRGRILLHNGAYSILTLQGLGISWLGRLSRLRELGSGLALIHNNTHLCFVHTV 480
DB 421 DLSVFQNLQVIRGRILLHNGAYSILTLQGLGISWLGRLSRLRELGSGLALIHNNTHLCFVHTV 480
QY 481 PWDQLFRNPHQALLHTANRPEDECVGEGGLACHQLCARGHCWGPPTQCVCNCSQFLRGQEC 540
DB 481 PWDQLFRNPHQALLHTANRPEDECVGEGGLACHQLCARGHCWGPPTQCVCNCSQFLRGQEC 540
QY 541 VEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADOCVACAHYKOPPCVAVC 600
DB 541 VEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADOCVACAHYKOPPCVAVC 600
QY 601 PSGVKPDLSTYMPIWKFPDEEGACQPCPNCTHSCVDLDDKGPAPAEQASPLTISIVSAVVG 660
DB 601 PSGVKPDLSTYMPIWKFPDEEGACQPCPNCTHSCVDLDDKGPAPAEQASPLTISIVSAVVG 660
QY 661 ILLVVVLGVVFGILIKRROQKIRKVTMRLLQETELVEPLTPSGAMPNQAQRILKETEL 720
DB 661 ILLVVVLGVVFGILIKRROQKIRKVTMRLLQETELVEPLTPSGAMPNQAQRILKETEL 720
QY 721 RKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKELDEAYMAGVGP 780
DB 721 RKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKELDEAYMAGVGP 780
QY 781 YVSRLLGICLTSTVQLVQLMPYGCCLLDHVRNRRGLSGDQLLNWCQIAKMSYLEDVR 840
DB 781 YVSRLLGICLTSTVQLVQLMPYGCCLLDHVRNRRGLSGDQLLNWCQIAKMSYLEDVR 840
QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMWMALESILRRFT 900

|||||
841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEYHADGGKVPKIMMALESILRRFT 900
901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVWKWM 960
901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVWKWM 960
961 IDSECRPRFRELVSERFARMARDPQFVVIQNEIDLGPASPLDSTFYRSLLEDDMDGLVDA 1020
961 IDSECRPRFRELVSERFARMARDPQFVVIQNEIDLGPASPLDSTFYRSLLEDDMDGLVDA 1020
1021 EYLVPQQGFCDPAPGAGGVMHRRSSSTRSGGGDLTLGLEPSEEEAAPSPLAPSEG 1080
1021 EYLVPQQGFCDPAPGAGGVMHRRSSSTRSGGGDLTLGLEPSEEEAAPSPLAPSEG 1080
1081 AGSDVFGDGLGMAAKGLQSLPHDPSPLQRYSEDPTVPLSETDGVVAPLTCSPQPEYV 1140
1081 AGSDVFGDGLGMAAKGLQSLPHDPSPLQRYSEDPTVPLSETDGVVAPLTCSPQPEYV 1140
1141 NQPDVRPQPPSPREGPLPAAPAGATLERAKTLPKGNVGVKDVAFGGAVENPEYLTPO 1200
1141 NQPDVRPQPPSPREGPLPAAPAGATLERAKTLPKGNVGVKDVAFGGAVENPEYLTPO 1200
1201 GGAAPQHPHPPAFSPAFDNLVYWDQDPPPERGAPPESTFKGTPTAENPEYLGLDVPV 1255
1201 GGAAPQHPHPPAFSPAFDNLVYWDQDPPPERGAPPESTFKGTPTAENPEYLGLDVPV 1255

RESULT 15

AAR39568
ID AAR39568 standard; Protein; 1433 AA.

XX AC AAR39568;

XX DT 07-FEB-1994 (first entry)

XX DE Sequence of c-erbB-2 tumour antigen.

XX KW Tumour antigen; c-erbB-2; glycoprotein.

XX OS Homo sapiens.

XX PN WO9316185-A.

XX PD 19-AUG-1993.

XX PF 05-FEB-1993; 93WO-US01055.

XX PR 06-FEB-1992; 92US-0831967.

XX PA (CETU) CETUS ONCOLOGY CORP.

XX PA (CREA-) CREATIVE BIOMOLECULES INC.

XX PI Houston LL, Huston JS, Oppermann H, Ring DB;

XX DR WPI; 1993-272889/34.

XX DR N-PSDB; AAQ46083.

XX PT New single chain Fv polypeptide binding to C-erbB-2 tumour

XX PT antigen - for imaging or treating breast or ovarian cancer etc.

XX PS Disclosure; pages 48-54; 87pp; English.

XX CC c-erbB-2 refers to a protein antigen expressed on the surface of

XX CC tumour cells. such as breast and ovarian tumour cells, which is an

XX CC approx. 200,000 mol. wt. acidic glycoprotein having an isoelectric

XX CC pt. of about 5.3 (see AAQ46083, AAR39568). X in AAR39583 represents

XX CC the location of a stop codon in AAQ46083.

XX SQ Sequence 1433 AA;

Query Match 98.0%; Score 6655; DB 14; Length 1433;

Best Local Similarity 98.2%; Pred. No. 0;

Matches 1233; Conservative 3; Mismatches 19; Indels 0; Gaps 0;
QY 1 MELAALCRGLLALLAPPGAASQVCTGTDMLRLPASPTHLDMLRHLYQGCQVQGNL 60
DB 1 MELAALCRGLLALLAPPGAASQVCTGTDMLRLPASPTHLDMLRHLYQGCQVQGNL 60
QY 61 ELYTLPTNASLPLQDIQEVQGVLIHAHQVQVPLQRLIRVGTQLQFEDNYALAVLDNG 120
DB 61 ELYTLPTNASLPLQDIQEVQGVLIHAHQVQVPLQRLIRVGTQLQFEDNYALAVLDNG 120
QY 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDTILWKDIFPHKNQLA 180
DB 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDTILWKDIFPHKNQLA 180
QY 181 LTLIDNTRACHPCSPMCKGRCWGESSEDCOSLRTVCAGGCARCKGPLQYIKANSK 240
DB 181 LTLIDNTRACHPCSPMCKGRCWGESSEDCOSLRTVCAGGCARCKGPLQYIKANSK 240
QY 241 FIGITELKHSDDLACILHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
DB 241 AGCTGPKHSDCLACILHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
QY 301 YNYLSTDVSGSCTLVCPHMQEVTAEQTCRCKSKPCARVCYGLGMEHLREVRVTSAN 360
DB 301 YNYLSTDVSGSCTLVCPHMQEVTAEQTCRCKSKPCARVCYGLGMEHLREVRVTSAN 360
QY 361 IQEFAGCKITFGLSLAFLPESFDGDPASNTAPLOPELOVFPETLEEITGVLYISAWPDSL 420
DB 361 IQEFAGCKITFGLSLAFLPESFDGDPASNTAPLOPELOVFPETLEEITGVLYISAWPDSL 420
QY 421 DLSVFQNLQVIRGRILHNGAYSILTLQGLGTSWGLSLRLSRELGSGLALIHNTHLFCFVHT 480
DB 421 DLSVFQNLQVIRGRILHNGAYSILTLQGLGTSWGLSLRLSRELGSGLALIHNTHLFCFVHT 480
QY 481 PMDQLFRNPHQALLHTANRPEDECVGEGLACHOLCARGHCWGPGTQCVCNCSOFLRGQEC 540
DB 481 PMDQLFRNPHQALLHTANRPEDECVGEGLACHOLCARGHCWGPGTQCVCNCSOFLRGQEC 540
QY 541 VEECRVLOGLPREYVNAHCLPCHPECQPNQSVTCFGEADOCVACAHYKDPFCVARC 600
DB 541 VEECRVLOGLPREYVNAHCLPCHPECQPNQSVTCFGEADOCVACAHYKDPFCVARC 600
QY 601 PSGVKPDLSPYMPITWPFDEEGACQPCINCTHSCVDLDDKGCPEARASPLTSISAVVG 660
DB 601 PSGVKPDLSPYMPITWPFDEEGACQPCINCTHSCVDLDDKGCPEARASPLTSISAVVG 660
QY 661 ILLVVVLGVVFGILIKRROOKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
DB 661 ILLVVVLGVVFGILIKRROOKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
QY 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKIILDEAYMAGVGP 780
DB 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKIILDEAYMAGVGP 780
QY 781 YVSRLLGICLTSTVQLVTQMLPYGCLLDHVRENRGLSGQDLNLCWQIAKGNYSILEDVR 840
DB 781 YVSRLLGICLTSTVQLVTQMLPYGCLLDHVRENRGLSGQDLNLCWQIAKGNYSILEDVR 840
QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEYHADGGKVPKIMMALESILRRFT 900
DB 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEYHADGGKVPKIMMALESILRRFT 900
QY 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVWKWM 960
DB 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVWKWM 960
QY 961 IDSECRPRFRELVSERFARMARDPQFVVIQNEIDLGPASPLDSTFYRSLLEDDMDGLVDA 1020
DB 961 IDSECRPRFRELVSERFARMARDPQFVVIQNEIDLGPASPLDSTFYRSLLEDDMDGLVDA 1020
QY 1021 EYLVPQQGFCDPAPGAGGVMHRRSSSTRSGGGDLTLGLEPSEEEAAPSPLAPSEG 1080
DB 1021 EYLVPQQGFCDPAPGAGGVMHRRSSSTRSGGGDLTLGLEPSEEEAAPSPLAPSEG 1080

Qy	1081	AGSDVFDGDLGMGAAGLQSLPTHDPSPLOQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV	1140
Db	1081	.AGSDVFDGDLGMGAAGLQSLPTHDPSPLOQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV	1140
Qy	1141	NQPDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKOVFAFGGAVENPEYLTPO	1200
Db	1141	NQPDVRPQPPSPREGPLPAARPAAGATLERPKTILSPGKNGVVKOVFAFGGAVENPEYLTPO	1200
Qy	1201	GGAAPQHPHPPAFSPAFDNLYYWDQDPPPERGAPPSFKGTPTAENPEYLGLDVVP	1255
Db	1201	GGAAPQHPHPPAFSPAFDNLYYWDQDPPPERGAPPSFKGTPTAENPEYLGLDVVP	1255

Search completed: July 22, 2003, 08:40:50
 Job time : 43.9774 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:25:54 ; Search time 22.9062 Seconds
(without alignments)
5267.077 Million cell updates/sec

Title: SEQ4-325-339-12
Perfect score: 6814
Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYGLDVPV 1255
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 73.*
1: piri.*
2: piri2.*
3: piri3.*
4: piri4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6728	98.7	1255	1 A24571	protein-tyrosine k
2	5931	87.0	1260	1 TVRTNU	protein-tyrosine k
3	5921.5	86.9	1254	2 I48161	p-185 precursor -
4	3146	46.2	1210	1 GQHUE	epidermal growth f
5	3117	45.7	1210	2 A53183	epidermal growth f
6	3094.5	45.4	1223	1 TVCHLV	epidermal growth f
7	2966.5	43.5	1308	2 A47253	epidermal growth f
8	2669	39.2	1166	1 S06142	protein-tyrosine k
9	2414.5	35.4	1342	2 A36223	kinase-related tra
10	2328.5	34.2	1339	2 JC4367	epidermal growth f
11	1766.5	25.9	698	1 TVFVLV	protein-tyrosine k
12	1703	25.0	604	1 TVYUHV	protein-tyrosine k
13	1647	24.2	544	2 S35745	protein-tyrosine k
14	1640	24.1	545	2 S00727	kinase-related tra
15	1630.5	23.9	1330	1 GQFFE	epidermal growth f
16	1623	23.8	540	2 B44776	protein-tyrosine k
17	1621	23.8	540	1 TVFVEB	protein-tyrosine k
18	1509	22.1	644	2 A36325	epidermal growth f
19	1290	18.9	1323	2 E88257	protein let-23 [im
20	1290	18.9	1374	2 S70712	protein-tyrosine k
21	1204	17.7	1369	2 S70713	protein-tyrosine k
22	1162	17.1	1717	1 A45558	epidermal growth f
23	1126	16.5	527	2 A42032	epidermal growth f
24	975.5	14.3	843	2 A27131	epidermal growth f
25	806.5	11.8	346	2 S13807	protein-tyrosine k
26	754.5	11.1	311	2 S13808	protein-tyrosine k
27	727	10.7	1363	2 T43220	insulin-like growth
28	697	10.2	1382	1 INHUR	insulin receptor p
29	696	10.2	1383	2 A36080	insulin receptor p

ALIGNMENTS

RESULT 1

A24571
protein-tyrosine kinase (EC 2.7.1.112) erbB2 precursor - human
N;Alternate names: c-erb-B-2 protein precursor; kinase-related transforming protein erb
C;Species: Homo sapiens (man)
C;Date: 25-Oct-1987 #sequence revision 06-Dec-1996 #text change 11-Jun-1999
C;Accession: A24571; A25491; A4188; B44188; I59509; I57622
R;Yamamoto, T.; Ikawa, S.; Akiyama, T.; Semba, K.; Nomura, N.; Miyajima, N.; Saito, T.;
Nature 319, 230-234, 1986
A;Title: Similarity of protein encoded by the human c-erb-B-2 gene to epidermal growth
A;Reference number: A24571; MUID:86118663; PMID:2999974
A;Accession: A24571
A;Molecule type: mRNA
A;Residues: 1-1255 <YAM>
A;Cross-references: GB:K03363; NID:g31197; PIDN:CAA27060.1; PID:g31198
R;Semba, K.; Kamata, N.; Toyoshima, K.; Yamamoto, T.
Proc. Natl. Acad. Sci. U.S.A. 82, 6497-6501, 1985
A;Title: A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB-1/epider
A;Reference number: A25491; MUID:86016729; PMID:2995967
A;Accession: A25491
A;Molecule type: DNA
A;Residues: 737-1031 <SEM>
A;Cross-references: GB:M11767; NID:g182163; PIDN:AAA35808.1; PID:g553282
R;Cousens, L.; Yang-Feng, T.L.; Liao, Y.C.; Chen, E.; Gray, A.; McGrath, J.; Seeburg,
Science 230, 1132-1139, 1985
A;Title: Tyrosine kinase receptor with extensive homology to EGF receptor shares chromo
A;Reference number: A44188; MUID:86070181; PMID:2999974
A;Accession: A44188
A;Molecule type: DNA
A;Residues: 740-910 <COU1>
A;Cross-references: GB:M12036; NID:g183988; PIDN:AAA35978.1; PID:g183989
A;Accession: B44188
A;Molecule type: mRNA
A;Residues: 1-517, 'RALL', 522, 'S', 524-654, 'V', 656-1169, 'A', 1171-1255 <COU2>
A;Cross-references: GB:M11730; NID:g183986
R;King, C.R.; Kraus, M.H.; Aaronson, S.A.
Science 229, 974-976, 1985
A;Title: Amplification of a novel v-erbB-related gene in a human mammary carcinoma.
A;Reference number: I59509; MUID:85272597; PMID:2992089
A;Accession: I59509
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 832-909 <REX>
A;Cross-references: GB:I29395; NID:g459807; PIDN:AAA35809.1; PID:g459808
R;Tal, M.; King, C.R.; Kraus, M.H.; Ullrich, A.; Schlessinger, J.; Givol, D.
Mol. Cell. Biol. 7, 2597-2601, 1987
A;Title: Human HER2 (neu) promoter: evidence for multiple mechanisms for transcriptiona
A;Reference number: I57622; MUID:87286898; PMID:3039351
A;Accession: I57622
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-191 <TAL>

insulin receptor p
insulin-like growth
protein-tyrosine k
insulin receptor-r
insulin receptor-r
insulin-like growth
insulin-like growth
insulin receptor -
insulin receptor -
insulin receptor -
protein-tyrosine k
protein-tyrosine k
protein-tyrosine k
tyrosine kinase Mp
protein-tyrosine k
protein-tyrosine k

A;Cross-references: GB:M16792; NID:G183983; PIDN:AAA58637.1; PID:G553332
C;Comment: Amplification and overexpression of this erbB-related gene occurs in about 30
C;Genetics:
A;Gene: GDB:ERBB2; NGL; NEU; HER-2
A;Cross-references: GDB:120613; OMIM:164870
A;Map position: 17q21.1-17q21.1
A;Introns: 25/1; 75/3; 147/1; 883/3
A;Note: the list of introns is incomplete
C;Function:
A;Description: catalyzes the phosphorylation of a peptidyl tyrosine residue by ATP
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho
inase
F;1-21/Domain: signal sequence #status predicted <SIG>
F;22-1255/Product: protein-tyrosine kinase erbB2 #status predicted <MAT>
F;22-653/Domain: extracellular #status predicted <EXT>
F;70-304/Domain: EGF receptor extracellular domain repeat <BE1>
F;395-605/Domain: EGF receptor extracellular domain repeat <BE2>
F;654-675/Domain: transmembrane #status predicted <TM>
F;676-1255/Domain: intracellular #status predicted <INT>
F;718-983/Domain: protein kinase homology <KIN>
F;726-734/Region: protein kinase ATP-binding motif
F;68,124,187,259,530,571,629/Binding site: carbohydrate (Asn) (covalent) #status predict
F;686/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F;753/Active site: Lys #status predicted
F;1139,1221,1222,1248/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation)

Query Match 98.7%; Score 6728; DB 1; Length 1255;
Best Local Similarity 98.6%; Pred. No. 5.6e-267;
Matches 1238; Conservative 5; Mismatches 12; Indels 0; Gaps 0;

QY 1 MELAALCRWGLLALLPPGAASCTVCTGDMKRLPASPEHLDMRLHYGCGOVVGNL 60
DB 1 MELAALCRWGLLALLPPGAASCTVCTGDMKRLPASPEHLDMRLHYGCGOVVGNL 60
QY 61 ELTYLPTNASLFLQDIQEVGYVLI AHNOVRQVPLQRLIRVGTQLFEDNYALAVLDNG 120
DB 61 ELTYLPTNASLFLQDIQEVGYVLI AHNOVRQVPLQRLIRVGTQLFEDNYALAVLDNG 120
QY 121 DPLNNTPTVTCASGGI RELQRLSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
DB 121 DPLNNTPTVTCASGGI RELQRLSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180
QY 181 LTLIDTNRSRACHPCSPMKSGRCWGESSEDCQSLTRTVCAAGCARCKGPLPTDCCHQC 240
DB 181 LTLIDTNRSRACHPCSPMKSGRCWGESSEDCQSLTRTVCAAGCARCKGPLPTDCCHQC 240
QY 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPEGRTYFGASCVTACP 300
DB 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPEGRTYFGASCVTACP 300
QY 301 YNYLSTDVGSCTLCPLHNOVTHAEDGTQRCCKSPCARVCYGLGMOYIKANSKFTGIT 360
DB 301 YNYLSTDVGSCTLCPLHNOVTHAEDGTQRCCKSPCARVCYGLGMOYIKANSKFTGIT 360
QY 361 ELEPAGCKIFGSLAFIPESPDGPASNTAPLQPEQLQVETLEITGLYVISAWPDSLP 420
DB 361 IQEPAGCKIFGSLAFIPESPDGPASNTAPLQPEQLQVETLEITGLYVISAWPDSLP 420
QY 421 DLSVFQNLQVIRGILHNGAYSLTQGLIGISWGLRLSRLGSLALIHNNTHLCFVHTV 480
DB 421 DLSVFQNLQVIRGILHNGAYSLTQGLIGISWGLRLSRLGSLALIHNNTHLCFVHTV 480
QY 481 PWDQLFRPHQALLHTANRDEDCVGGGLACHQICARGHCWGPGPTQCVNCSQPLRQEC 540
DB 481 PWDQLFRPHQALLHTANRDEDCVGGGLACHQICARGHCWGPGPTQCVNCSQPLRQEC 540
QY 541 VEECRVLQGLPREVYNARHCLPCHPECPQNGSVTCFGRPADQCACAHYKDPFPCVARC 600
DB 541 VEECRVLQGLPREVYNARHCLPCHPECPQNGSVTCFGRPADQCACAHYKDPFPCVARC 600
QY 601 PSGVKPDLSYMPIWKFDEEGACQCPINCTHSCVDLDDKCPAEORASPLTISVAVVG 660
DB 601 PSGVKPDLSYMPIWKFDEEGACQCPINCTHSCVDLDDKCPAEORASPLTISVAVVG 660

DB 601 PSGVKPDLSYMPIWKFDEEGACQCPINCTHSCVDLDDKCPAEORASPLTISVAVVG 660
QY 661 ILLVVVLGVVFGILIKRRQOKIRKYIMRRLLOTELVEPLTPSGAMPNQAMRILKTEL 720
DB 661 ILLVVVLGVVFGILIKRRQOKIRKYIMRRLLOTELVEPLTPSGAMPNQAMRILKTEL 720
QY 721 RKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILD EAYVMAGVGP 780
DB 721 RKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILD EAYVMAGVGP 780
QY 781 YVSRLLGICLTSTVQLVTLQMPYVCLLDHVRENRGLGSDLLNWCQIAKGSYLEVDV 840
DB 781 YVSRLLGICLTSTVQLVTLQMPYVCLLDHVRENRGLGSDLLNWCQIAKGSYLEVDV 840
QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMWALESLIRRRFT 900
DB 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMWALESLIRRRFT 900
QY 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYIMVKCWM 960
DB 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYIMVKCWM 960
QY 961 IDSECRPRFRELVSFERSMARDPQRFVVIQNEGLGASPLDSTFYRSLLEDDDMGLVDA 1020
DB 961 IDSECRPRFRELVSFERSMARDPQRFVVIQNEGLGASPLDSTFYRSLLEDDDMGLVDA 1020
QY 1021 EYLVLPQOQFFCDDPAPGAGGMVHHRSSSTSGGGDLTLGLEPSEEAAPRSLAPSEG 1080
DB 1021 EYLVLPQOQFFCDDPAPGAGGMVHHRSSSTSGGGDLTLGLEPSEEAAPRSLAPSEG 1080
QY 1081 AGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVVAPLTCSPQPEYV 1140
DB 1081 AGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVVAPLTCSPQPEYV 1140
QY 1141 NQPDVPPQPPSPREGPLPAARPAATLERAKTLSPGKNGVVDVAFAGGAVENPEYLTPO 1200
DB 1141 NQPDVPPQPPSPREGPLPAARPAATLERAKTLSPGKNGVVDVAFAGGAVENPEYLTPO 1200
QY 1201 GGAAPHPHPPAPSPAFDNLVYWDQPPERGAAPPSTFKGTPTAENPEYLGLDVVP 1255
DB 1201 GGAAPHPHPPAPSPAFDNLVYWDQPPERGAAPPSTFKGTPTAENPEYLGLDVVP 1255

RESULT 2
TVRINU
protein-tyrosine kinase (EC 2.7.1.112) neu precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 31-Dec-1998 #sequence_revision 31-Dec-1998 #text_change 11-Jun-1999
C;Accession: A24562; A61204
R;Bargmann, C.I.; Hung, M.C.; Weinberg, R.A.
Nature 319, 226-230, 1996
A;Title: the neu oncogene encodes an epidermal growth factor receptor-related protein.
A;Reference number: A24562; MUID:86118662; PMID:3945311
A;Accession: A24562
A;Molecule type: mRNA
A;Residues: 1-1260 <BAR>
A;Cross-references: EMBL:X03362; NID:G56745; PIDN:CAA27059.1; PID:G56746
R;Masui, T.; Mann, A.M.; Macatee, T.E.; Garland, E.M.; Okamura, T.; Smith, R.A.; Cohen,
Carcinogenesis 12, 1975-1978, 1991
A;Title: Direct DNA sequencing of the rat neu oncogene transmembrane domain reveals no
2-thiazolylformamide or N-methyl-N-nitrosourea.
A;Reference number: A61204; MUID:92035293; PMID:1682063
A;Accession: A61204
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 637-663, 'V', 665-702 <MAS>
A;Note: authors translated the codon GCA for residue 25 as Val
C;Genetics:
A;Gene: neu
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phosph
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-1260/Product: protein-tyrosine kinase neu #status predicted <MAT>

F:658-680/Domain: transmembrane #status predicted <TM>
F:723-988/Domain: protein kinase homology <KIN>
F:731-739/Region: protein kinase ATP-binding motif
F:71,191,263,535,576,634/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:691/Binding site: phosphate (Thr) (covalent) #status predicted
F:758/Active site: Lys #status predicted
F:882,1227,1253/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 87.0%; Score 5931; DB 1; Length 1260;
Best Local Similarity 87.0%; Pred. No. 1.6e-234;
Matches 1093; Conservative 52; Mismatches 110; Indels 2; Gaps 2;

QY 1 MELAALCRWGLLLALLPPGAASTQVCTGTDMLRLPASPTHLDMLRHLYQGCVVQGNL 60
DB 4 MELAAWCRWGLLLALLPPGIAGTQVCTGTDMLRLPASPTHLDMLRHLYQGCVVQGNL 63
QY 61 ELTYLPTNASLSFLQDIQEVQVYLIHNOVRQVPLQRLIRVGTQLFEDNYALAVLDNG 120
DB 64 ELTYVPAASLSFLQDIQEVQVYLIHNOVRQVPLQRLIRVGTQLFEDNYALAVLDNR 123
QY 121 DPLNNTTPVT-GASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDTILWKDIFHKNNQL 179
DB 124 DPQDNVAASPGRTEGLRELQRLSLTEILKGGVLIQORNPOLCYQDMVLWKDIFHKNNQL 183
QY 180 ALTLIDNRSRACHPCSPCKGSRGWESSEDCOSLRTVTCAGCARCKGPLTDCCHEQ 239
DB 184 APVDIDNRSRACHPCSPCKGSRGWESSEDCQILTTCTGCGARCKGRLPTDCCHEQ 243
QY 240 CAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRTFGASCVTAC 299
DB 244 CAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRTFGASCVTTC 303
QY 300 PYNLSTDVGSCTLVCPNNQEVTAEDGTQRCCKSKPCARVCYGLGMQVIAKNSKFIGI 359
DB 304 PYNLSTEVGSCTLVCPNNQEVTAEDGTQRCCKSKPCARVCYGLGMHLRGARITSD 363
QY 360 TELEFAGCKKI FGSFLAPLPSFGDPSASNTAPIQPEOLQVFTEILTEITGLYVISAHPDSL 419
DB 364 NVQEFDPCKKI FGSFLAPLPSFGDPSASNTAPIQPEOLQVFTEILTEITGLYVISAHPDSL 423
QY 420 PDLVSFQNLQVIRGILHNGAYSILTLQGLIGISWGLRLSRELGSGLAIHNNHLCFVHT 479
DB 424 RDLVSFQNLRIIRGILHNGAYSILTLQGLIGISWGLRLSRELGSGLAIHNNHLCFVHT 483
QY 480 VPMDQLFRNPQALLHTANRPEDE-CVGEGLACHQLCARGHCWGPQTCVNCQSFLRGQ 538
DB 484 VPMDQLFRNPQALLHSGNRPEDLCVSSGLVCNLSLACHGCHWGPQTCVNCSHFLRGQ 543
QY 539 ECVEECRVGLQGLPREYVNAHCLPCHPECOPOQNGSVTCFQPEADQCVACHYKDPDFCVA 598
DB 544 ECVEECRVWKGKLPREYVSDKRLPCHPECOPOQNSSETCFGSEADQCAACHYKDSSCVA 603
QY 599 RCPGKVKPDLSPYMPINWKFPEEGACOPCINCTHSCVDLDDKCPAEORASPLTSIVSAV 658
DB 604 RCPGKVKPDLSPYMPINWKFPEEGICOPCINCTHSCVDLDDERCPAEORASPLTIVATV 663
QY 659 VGILLVVGLVFWGILIKRRQKIRKYTMERLLQETELVEPLTPSGAMPNQAMRILKET 718
DB 664 EGVLFLIILVVVGLIKRRQKIRKYTMERLLQETELVEPLTPSGAMPNQAMRILKET 723
QY 719 ELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVAGVG 778
DB 724 ELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVAGVG 783
QY 779 SPVVSRLIGLCLTSTVOLTPMPYGCCLLDHVENRGLSGODLLNCKMCIQAKGMSYLED 838
DB 784 SPVVSRLIGLCLTSTVOLTPMPYGCCLLDHVENRGLSGODLLNCKMCIQAKGMSYLED 843
QY 839 VRLVHRDLAARNVNLKSPNPKVITDFGLARLLDIDETEHADGKGKVPKMMALLESILRRR 898
DB 844 VRLVHRDLAARNVNLKSPNPKVITDFGLARLLDIDETEHADGKGKVPKMMALLESILRRR 903
QY 899 FTHQSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPPICTIDVYIMVVK 958

DB 904 FTHQSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPPICTIDVYIMVVK 963
QY 959 WMIDSECRPRFRELVESEFSEFARMARDPQRFVVIQNEIDLGSPASPLDSTFYRSILLEDDMGDLV 1018
DB 964 WMIDSECRPRFRELVESEFSEFARMARDPQRFVVIQNEIDLGSPSPMDSTFYRSILLEDDMGDLV 1023
QY 1019 DAEYLVPOQGFPCPDPAFGAGGMVHRRHSSTRSGGDLTLGLPSESEEAAPRSLAPS 1078
DB 1024 DAEYLVPOQGFPCPDPTCTGTAHRRHSSTRSGGDLTLGLPSESEEAAPRSLAPS 1083
QY 1079 EGAGSDVDFGDLGMAAGKGLQSLPTHDPLOYSEDPTVPLPSETDGYVAPITCPOPE 1138
DB 1084 EGAGSDVDFGDLGMAAGKGLQSLPTHDPLOYSEDPTVPLPSETDGYVAPITCPOPE 1143
QY 1139 YVNPQDVRPQPPSPREGPLPAARPGATLERAKTLPSPKNGVVKDVFAFGGAVENPEYLT 1198
DB 1144 YVNOSEVQPPPLTPGGLPPVPRPGATLERAKTLPSPKNGVVKDVFAFGGAVENPEYLV 1203
QY 1199 POGGAAPQHPPPAFSPAFDNLVYWDQDPPERGAPESTFKGTPTAENPEYLGLDVVP 1255
DB 1204 PREGTASPHPSPAFSPAFDNLVYWDQDPPERGAPESTFKGTPTAENPEYLGLDVVP 1260

RESULT 3
I48161
p-185 precursor - golden hamster
C:Species: Mesocricetus auratus (golden hamster)
C>Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 18-Jun-1999
C:Accession: I48161
R:Nakamura, T.; Ushijima, T.; Ishizaka, Y.; Nagao, M.; Arai, M.; Yamazaki, Y.; Ishikawa
Gene 140, 251-255, 1994
A:Title: Cloning and activation of the Syrian hamster neu proto-oncogene.
A:Reference number: I48161; MUID:94193007; PMID:7908275
A:Accession: I48161
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1254 <RES>
A:Cross-references: GB:D16295; NID:9493236; PID:BAA03801.1; PID:g747595
C:Genetics:
A:Gene: neu
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP
F:718-983/Domain: protein kinase homology <KIN>
F:726-734/Region: protein kinase ATP-binding motif

Query Match 86.9%; Score 5921.5; DB 2; Length 1254;
Best Local Similarity 86.6%; Pred. No. 3.9e-234;
Matches 1087; Conservative 61; Mismatches 106; Indels 1; Gaps 1;

QY 1 MELAALCRWGLLLALLPPGAASTQVCTGTDMLRLPASPTHLDMLRHLYQGCVVQGNL 60
DB 1 MELAAWCRWGLLLALLPPGAGTQVCTGTDMLRLPASPTHLDIVRHLYQGCVVQGNL 60
QY 61 ELTYLPTNASLSFLQDIQEVQVYLIHNOVRQVPLQRLIRVGTQLFEDNYALAVLDNG 120
DB 61 ELTYLPAANATLSFLQDIQEVQVYLIHNOVRQVPLQRLIRVGTQLFEDNYALAVLDNR 120
QY 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDTILWKDIFHKNNQLA 180
DB 121 DPLDNVTATGTATPGLRELQRLSLTEILKGGVLIQORNPOLCYQDTILWKDIFHKNNQLA 180
QY 181 LTLIDNRSRACHPCSPCKGSRGWESSEDCOSLRTVTCAGCARCKGPLTDCCHEQ 240
DB 181 PVDIDNRSRACHPCSPCKGSRGWESSEDCQTLTGTIAPRAVPAARARLPTDCCHEQ 240
QY 241 AAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRTFGASCVTAC 300
DB 241 AAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRTFGASCVTTC 300
QY 301 PYNLSTDVGSCTLVCPNNQEVTAEDGTQRCCKSKPCARVCYGLGMQVIAKNSKFIGIT 360
DB 301 PYNLSTEVGSCTLVCPNNQEVTAEDGTQRCCKSKPCARVCYGLGMHLRGARITSAN 360

Qy 361 ELBPAGCKIFGSLAFIPESFDGDPASNTAPLOPEQLQVFETLEEITGYLYISAWPDSL 420
Db 361 IQEPAGCKIFGSLAFIPESFDGNPSSGIAPLTPPEQLQVFETLEEITGYLYISAWPDSL 420
Qy 421 DLSVFQNLQVIRGILHNGAYSITLOGLGISWGLRSLRGLSLALIHNTHLFCFVHTV 480
Db 421 DLSVFQNLQVIRGILHNGAYSITLOGLGISWGLRSLRGLSLALIHNTHLFCFVHTV 480
Qy 481 PWDQLFRNPQALLHTANRDEDECVEGGLACHQLCARGHCWGPGPTQCVNCSFPLRQEC 540
Db 481 PWDQLFRNPQALLHSGNPSEECGLKDFACYPICAHGHCWGPGPTQCVNCSFPLRQEC 540
Qy 541 VEECRVLQGLPREVVNARHCLPCHPECPQNGSVTCFGEADQVCAHYKDPFFCVARC 600
Db 541 VKECRVWGLPREVVNARHCLPCHPECPQNGSVTCFGEADQVCAHYKDPFFCVARC 600
Qy 601 PSQVKPDLNYPINWKFDEEGACQPCPINCCHSCVDLDDKCPAEQQRASPLTSTVSAVG 660
Db 601 PSQVKPDLNYPINWKFDEEGACQPCPINCCHSCVDLDDKCPAEQQRASPLTSTVSAVG 660
Qy 661 ILLVVVGVLGILIKRQKIRKYTWRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720
Db 661 ILLFLVIGVVVGLIKRQKIRKYTWRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720
Qy 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGVGP 780
Db 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGVGP 780
Qy 781 YVSRLLGICLTSTVQLTQMPYGCLLDHVRENRGRGLGSDLLNWCQIAKGSYLEDVR 840
Db 781 YVSRLLGICLTSTVQLTQMPYGCLLDHVRENRGRGLGSDLLNWCQIAKGSYLEDVR 840
Qy 841 LVHRDLAARNVLKSPNHVKITDGLARLLDIDETEHADGKVPKIKMALESILRRRFT 900
Db 841 LVHRDLAARNVLKSPNHVKITDGLARLLDIDETEHADGKVPKIKMALESILRRRFT 900
Qy 901 HQSDVWSVGVTVWELMTFGKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYIMVWKCM 960
Db 901 HQSDVWSVGVTVWELMTFGKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYIMVWKCM 960
Qy 961 IDSECRPRFELVSEFSEMRARDPQFVVIQNEIDLGPASPLDSTFYRSLLEDGDDGLVDA 1020
Db 961 IDSECRPRFELVSEFSEMRARDPQFVVIQNEIDLGPASPLDSTFYRSLLEDGDDGLVDA 1020
Qy 1021 EYLVPOQGFPPDPAPCAGSMVHRHRSSTRSGGDLTLGLPSEEEAPRSPAPSEG 1080
Db 1021 EYLVPOQGFPPDPAPCAGSMVHRHRSSTRSGGDLTLGLPSEEEAPRSPAPSEG 1080
Qy 1081 AGSDVFDGLGMAKGLQSLPTHPDPLQRYSEDPTVPLPSETDGVVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGLGMAKGLQSLPTHPDPLQRYSEDPTVPLPSETDGVVAPLTCSPQPEYV 1140
Qy 1141 NQDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKOVFAFGGAVENPEYLTPQ 1200
Db 1141 NQDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKOVFAFGGAVENPEYLTPQ 1200
Qy 1201 GGAAPQHPHPPAFSPADNLVYWDQDPPERGAPSTFKGPTTANPEYLGLDVVP 1255
Db 1201 GGSASQPH-PPALCPAFDNLVYWDQDPPERSGPPNTEGPTAENPEYLGLDVVP 1254

RESULT 4

GQHUE

N;Contains: epidermal growth factor precursor - human

C;Species: Homo sapiens (man)

C;Date: 15-Nov-1984 #sequence revision 27-Nov-1985 #text change 11-Jun-1999

C;Accession: A00641; A25772; S30024; A38672; A00642; A43615; A23062; A05281; A33

R;Ullrich, A.; Coussens, L.; Hayflick, J.S.; Dull, T.J.; Gray, A.; Tam, A.W.; Lee, J.; Y

rg, P.H.

A;Title: Human epidermal growth factor receptor cDNA sequence and aberrant expression of

A;Reference number: A00641; MUID:84219729; PMID:6328312
A;Accession: A00641
A;Molecule type: mRNA
A;Residues: 1-1210
A;Cross-references: EMBL:X00588; NID:g31113; PIDN:CAA25240.1; PID:g575924
A;Note: the authors translated the codon AAG for residue 540 as Asn
R;Shii, S.; Xu, Y.; Stratton, R.H.; Roe, B.A.; Merlino, G.T.; Pastan, I.
Proc. Natl. Acad. Sci. U.S.A. 82, 4920-4924, 1985
A;Title: Characterization and sequence of the promoter region of the human epidermal gr
A;Reference number: A25772; MUID:85270438; PMID:2391899
A;Accession: A25772
A;Status: translation not shown
A;Molecule type: DNA
A;Residues: 1-29 <ISH>
A;Cross-references: GB:M11234; NID:gl81981; PIDN:AAA52370.1; PID:g553272
R;Haley, J.; Whittle, N.; Bennett, P.; Kinchington, D.; Ullrich, A.; Waterfield, M.
Oncogene Res. 1, 375-396, 1987
A;Title: The human EGF receptor gene: structure of the 110 kb locus and identification
A;Reference number: S30024; MUID:88217333; PMID:3329716
A;Accession: S30024
A;Molecule type: DNA
A;Residues: 1-29 <HA2>
A;Cross-references: EMBL:X06370; NID:g31118; PIDN:CAA29668.1; PID:g31119
R;Haley, J.D.; Waterfield, M.D.
J. Biol. Chem. 266, 1746-1753, 1991
A;Title: Contributory effects of de Novo transcription and premature transcript termina
A;Reference number: A38672; MUID:91107677; PMID:1988448
A;Accession: A38672
A;Molecule type: DNA
A;Residues: 1-29 <HAL>
A;Cross-references: GB:M38425; NID:gl81977; PIDN:AAA63171.1; PID:g553271
A;Experimental source: carcinoma cell line A431-7
R;Xu, Y.; Ishii, S.; Clark, A.J.L.; Sullivan, M.; Wilson, R.K.; Ma, D.P.; Roe, B.A.; Me
Nature 309, 806-810, 1984
A;Title: Human epidermal growth factor receptor cDNA is homologous to a variety of RNAs
A;Reference number: A00642; MUID:84245835; PMID:6330563
A;Accession: A00642
A;Molecule type: mRNA
A;Residues: 150-187, 'KSVIQAV', 195, 'M', 197, 'A', 199-222, 'S', 224-304, 'RA', 307-32
'798-799, 'TD', 802-811, 'R', 813-942 <XUY>
A;Experimental source: A431 human carcinoma cells, which have large numbers of EGF rece
R;Lin, C.R.; Chen, W.S.; Krueger, W.; Stolarsky, L.S.; Weber, W.; Evans, R.M.; Verma, I
Science 224, 843-848, 1984
A;Title: Expression cloning of human EGF receptor complementary DNA: gene amplification
A;Reference number: A43615; MUID:84196372; PMID:6326261
A;Accession: A43615
A;Molecule type: mRNA
A;Residues: 713-964 <LIN>
A;Experimental source: epidermoid carcinoma cell line A431
R;Simmen, F.A.; Gope, M.L.; Schulz, T.Z.; Wright, D.A.; Carpenter, G.; O'Malley, B.W.
Biochem. Biophys. Res. Commun. 124, 125-132, 1984
A;Reference number: A23062; MUID:85046483; PMID:6093780
A;Accession: A23062
A;Molecule type: mRNA
A;Residues: 1028-1210 <SIM>
R;Weber, W.; Gill, G.N.; Speiss, J.
Science 224, 294-297, 1984
A;Reference number: A05281; MUID:84172183; PMID:6324343
A;Accession: A05281
A;Molecule type: protein
A;Residues: 25-30, 'S', 32-51, 454-467 <WEB>
R;Russo, M.W.; Lukas, T.J.; Cohen, S.; Staros, J.V.
J. Biol. Chem. 260, 5205-5208, 1985
A;Title: Identification of residues in the nucleotide binding site of the epidermal gro
A;Reference number: A60143; MUID:85182650; PMID:2985580
A;Accession: A60143
A;Molecule type: protein
A;Residues: 740-744, 'X', 746-747 <RUS>
R;Mroczkowski, B.; Mosig, G.; Cohen, S.
Nature 309, 270-273, 1984
A;Title: ATP-stimulated interaction between epidermal growth factor receptor and superc
A;Reference number: A38023; MUID:84191554; PMID:6325948
A;Contents: annotation; receptor activity

A:Molecule type: mRNA
 A:Residues: 969-971, 'K', 973-1115, 'D' <EIS>
 A:Cross-references: EMBL:Z12608
 R:Heiseermann, G.J.; Gill, G.N.
 J. Biol. Chem. 263, 13152-13158, 1988
 A:Title: Epidermal growth factor receptor threonine and serine residues phosphorylated
 A:Reference number: A28941; MUID:88330814; PMID:3138233
 A:Accession: A28941
 A:Molecule type: protein
 A:Residues: 689-694, 'X', 696-704, 'L', 706-707, 989-992, 'XX', 995-996, 'X', 998-1000; 1002-1009,
 R:Hibbs, M.L.; Dunn, A.R.; Alexander, W.S.
 submitted to the EMBL Data Library, April 1994
 A:Description: The complete cDNA sequence of the Mouse Epidermal Growth Factor Receptor
 A:Reference number: S45325
 A:Accession: S45325
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-971, 'K', 973-1210 <VER>
 A:Cross-references: EMBL:X78987; NID:9488830; PIDN:CAA55587.1; PID:9488831
 R:Paria, B.C.; Das, S.K.; Andrews, G.K.; Dev, S.K.
 Proc. Natl. Acad. Sci. U.S.A. 90, 55-59, 1993
 A:Title: Expression of the epidermal growth factor receptor gene is regulated in mouse b
 A:Reference number: I49643; MUID:93126380; PMID:7678348
 A:Accession: I49643
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 12-20, 22-132 <RES>
 A:Cross-references: GB:L06864; NID:9193001; PIDN:AAA53029.1; PID:9567201
 C:Genetics:
 A:Gene: EGFR
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: ATP; growth factor receptor; kinase-related transforming protein; phosphop
 F:1-24/Domain: signal sequence #status predicted <SIG>
 F:648-670/Domain: transmembrane #status predicted <TM>
 F:712-977/Domain: protein kinase homology <KIN>
 F:720-728/Region: protein kinase ATP-binding motif
 F:680,695/Binding site: phosphate (Thr) (covalent) #status experimental
 F:697,1070,1071/Binding site: phosphate (Ser) (covalent) #status experimental
 F:993/Binding site: (or 997) phosphate (Ser) (covalent) #status experimental
 F:1028/Binding site: (or 1030 or 1032) phosphate (Ser) (covalent) #status experimental
 F:1197/Binding site: phosphate (Tyr) (covalent) #status experimental

Query Match 45.7%; Score 3117; DB 2; Length 1210;
 Best Local Similarity 49.8%; Pred. No. 6e-120;
 Matches 634; Conservative 168; Mismatches 359; Indels 112; Gaps 25;
 Qy 11 LLLALLPPGAA--STQVCTGDMKRLRLPASPTHLDMLRHLYQCGVQVQGNLEITYLPTN 68
 Db 14 LLLTALCAAGGAALEKKVCCQTSNRLTQLGTGFEDHFLSLQRMNCEVVLGNLEITYVQRN 73
 Qy 69 ASLSFLQDIOEVQGVYLAHQVRCVPLQRLRIVRGTOLFDNYALAVLDNGDPLNNTTP 128
 Db 74 YDLGFLKTIQEVAGVYLAHNTVERIPLNQLIQRGNALYENTYALISN----- 124
 Qy 129 VTGASPGRLQLRLSLTEILKGGVLIQRNQLCVQDTILWKDI----FHKNNALALTLI 184
 Db 125 -YGTNRTGURELPNRELQELIIGAVRFSNNPILCNWDIQRDIVQNVFMSNMDL---- 180
 Qy 185 DITNRSRACHPCSPMCKGRWCSESDCOSLTRTVACGCA-RCKGPLPTDCHEQCAAG 243
 Db 181 -QSHPSSCPCKDPSCPNGSCNGGGBEENCKLTKIICAQCCSHRCRGRSPDCHNQCAAG 239
 Qy 244 CTGPKHSDCLAHFNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACPNY 303
 Db 240 CTGPRESCLVCQKQFQDEATCKOTCPPLMLNPTTYQMDVNPPEGKYSFGATCVKCKPRNY 299
 Qy 304 LSTDVGSCTVLCPLHNOVQVTAEDGTORCEKSKPCARVCYGLGMQYIKANSKFTIGITELE 363
 Db 300 VVTDHGSVCVRAGGPDYIEV-EDGIRKCKKCDGCRKVCNGIGIGFEK-DTILSNATNIK 357
 Qy 364 -FAGCKITFGSLAFIPESFGDGPASNTAPLOEQLOVFEITGYLYISAWPDSPLDL 422
 Db 358 HFKYCTAISGDLHILPFAVKGDSFTRTPPLDPRLEILKTVKEITGFLIIQAWPDNWTDL 417

RESULT 6

TVCHLV

epidermal growth factor receptor precursor - chicken

N;Contents: protein-tyrosine kinase (EC 2.7.1.112) erbB

C;Species: Gallus gallus (chicken)

C;Date: 28-Feb-1986 #sequence_revision 05-May-1995 #text_change 04-Feb-2000

C;Accession: A27720; A00643

R;Lax, I.; Johnson, A.; Howk, R.; Sap, J.; Bellot, F.; Winkler, M.; Ullrich, A.; Vennet

Mol. Cell. Biol. 8, 1970-1978, 1988

A;Title: Chicken epidermal growth factor (EGF) receptor: cDNA cloning, expression in mo

A;Reference number: A27720; MUID:88261272; PMID:3260329

A;Accession: A27720
A;Molecule type: mRNA
A;Residues: 1-1223 <LAX>
A;Cross-references: GB:M20386
R;Nilsson, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M.
Cell 41, 719-726, 1985
A;Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro
A;Reference number: A00643; MUID:85228222; PMID:2988784
A;Accession: A00643
A;Molecule type: mRNA
A;Residues: 585-1223 <NTL>
A;Cross-references: GB:M10066
C;Genetics:
A;Gene: erbB
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: alternative splicing; ATP; autophosphorylation; glycoprotein; growth factor
specific protein kinase
F;1-30/Domain: signal sequence #status predicted <SIG>
F;31-1223/Product: epidermal growth factor receptor #status predicted <MAT>
F;31-654/Domain: extracellular #status predicted <EXT>
F;81-307/Domain: EGF receptor extracellular domain repeat <EB1>
F;397-610/Domain: EGF receptor extracellular domain repeat <EB2>
F;655-677/Domain: transmembrane #status predicted <TM>
F;678-1223/Domain: intracellular #status predicted <INT>
F;719-984/Domain: protein kinase homology <KIN>
F;727-735/Region: protein kinase ATP-binding motif
F;136,202,280,361,370,422,575,580,615,635/Binding site: carbohydrate (Thr) (covalent) #
F;192,650/Binding site: carbohydrate (Ser) (covalent) #status predicted
F;687/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
F;754/Active site: Lys #status predicted
F;1100,1183,1208/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #stat

Query Match 45.4%; Score 3094.5; DB 1; Length 1223;
Best Local Similarity 48.7%; Pred. No. 5e-119;
Matches 632; Conservative 173; Mismatches 346; Indels 147; Gaps 27;

QY 8 RWGLLLALLPPGAA-----STVCTGTDMKRLPASPETHDMLRHLHYGCGVQVGNLE 61
DB 13 RGAALVLLLLGVALCSAVEBEKKVCGQTNNKLTOLGHVEDHFTSLQRMVNNCEVLSNLE 72
QY 62 LTVLPTNASLSFLQDIOEVGYVLIHNOVQVLPQRLRVGTQLPEDNYALAVLNDG 121
DB 73 ITVEHNRDLTFLKTIQEVAGYVLIHNMVDVPLENLIQIRGNVLYDNSFALAVLSNYH 132
QY 122 PLNNTPTVGTASPGGLRELQSLRSLTEILKGGVLITQPNQCYQDTILWKDIFHKNQNAL 181
DB 133 -MNKTQ-----GLRELPMKRLSELINGGVKISNNPKLCNMDVTLWNIIIDTSRK-PL 182
QY 182 TLID-TNRSRACHPCSPMKGSRGWSSBDCSLTRTVCAAGCA-RCKGLPTDCCHEQ 239
DB 183 TVLDFASNLSSCPKCHPNECTEDHCWGAGEQNCQTLTKVICAQQCSCGRGKVPSCCHNQ 242
QY 240 CAAGCTGPKHSDCLACLFHNSHIGELHCPALVTYNTDTFESMPNPEGRYTFGASCVTAC 299
DB 243 CAAGCTGPRSDCLACKKFRDDATCKDTCPLVLYNPPTYOMDVNPGKYSFGATCVR 302
QY 300 PYNVLTSDVGSCTLVCPHNNQVETAEQTCRCKSCPCARVCVYGLGMQVIKANSKFIGI 359
DB 303 PHNVVYTDHSCVRSNCNTDIYEV-EENGVRKCKCDGLSKVCNGVIGIGELKGLS-INA 360
QY 360 TELE-PAGCKKIFGSLAFLESFGDGPASNTAPLOPQLOVFFTEBITGYLYISAWPDS 418
DB 361 TNIDSFNCKTNGIDVLSILVAFLDGAFKTLPLDPKKLDVFRVTKVSEISGFLLIQAWPDN 420
QY 419 LPDLSVFQNLQVTRGLHNGVSLTLQGLIGISWGLRSRLRSLGSLALHHTHLICFVH 478
DB 421 ATDLYAFENLEIRGTRKQHQVSLAVVNLKIOSLGRSLRSLKETSDGDIAMKNNCLCYAD 480
QY 479 TVPMDQLFRNPQALLHTANRPEDCVGEGLACHOLCARGHCWGPPTQCVCNCSQFLRGQ 538
DB 481 TMNWSLSFATQSQTKIIQNNKNDCTADRHVCDPLCSDVCGCGGPGFHCFCRFRSRQK 540
QY 539 ECVEECRVLQGLPREYVVARHCLFCHPECPQNG---SVTCFGPEADQCACAHYKDPFP 595

DB 541 ECVKQCNILQGEPRERDSDKCLPCHSECLVQNSTAYNTTCSGPGPDHCKCAHFIDGPH 600
QY 596 CVARCPGKVPDLISYMPIWKFPEDEGACPCPINCTHSCVDLDDKGCPCAPQASPLTSIV 655
DB 601 CVKACFAGVUGENDTL-VWKYADANAVCQCHPNCTRGCKGPGLEGCP---NGSKTSPSIA 656
QY 656 SAVV-GILLVVVILGVVFGILIKRRQOKIRKYTMRRLLQETELVEPLTPPGAMPNQAOMRI 714
DB 657 AGVVGGLLCLVVVGLGIGLYLRR-HIVKRTLRLLQERELVEPLTPSGEAPNQAHLRI 715
QY 715 LKETEARKVKVGLSGAFGTVYKGIWIPDGENVKIPIVAKVLRRENTSPKANKEILDEAYVM 774
DB 716 LKETEARKVKVGLSGAFGTVYKGIWIPDGENVKIPIVAKVLRRENTSPKANKEILDEAYVM 775
QY 775 AGVSPVSVRLIGLICLTSTVOLTPQMPYGCCLLDHYRENRGRGLSGDOLLNWCQIAKMS 834
DB 776 ASVDNPHVCCLLIGLICLTSTVOLITQMPYGCCLLDYIREHKDNIGSQVLLNWCQIAKMN 835
QY 835 YLEDVRLVHRDLAARNVVKSPNHVKITDFGLARLLDIDETEHADGKVPKIKWMALES 894
DB 836 YLSERLVRDLAARNVVKTPQHKITDFGLAKLLGADSEKEYHAEGKVPKIKWMALES 895
QY 895 LRRRPTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVMI 954
DB 896 LHRITHTQSDVMSYGVTVWELMTFGSKPYDGIIPASEISSVLEKGERLPQPPICTIDVMI 955
QY 955 MVKCMWIDSCRRFRRELYSESRMARDPQRFVVIQ-NEDLPASPLDSTFYSLLEDDDD 1013
DB 956 MVKCMWIDADSKFKRELTAEFKMDRPPRYLVIOGDERKMLPSPSTDKFYTLMBEED 1015
QY 1014 MGDVDAEYLVPOQGFPCDPAPAGGMVHHRSSSTRSGGDLTLGLEPSEEEAPRS 1073
DB 1016 MEDIVDAEYLVHQGF-----NSPST-----SRT 1041
QY 1074 PL-----APSEAGSDVFDGLGMAAKGLQSLPTHDPSPLOYRSDPTVPLPSET--DG 1126
DB 1042 PLLSSLSATSNNATNCID-----RNGQGHVREDSPVQRYSSDPTGNFLEESIDDG 1093
QY 1127 VYAPLTCSPQEVVNVQVDRPQPPSPREGPLPAARAGATLERAKTLSPKNGVGVKDFV- 1185
DB 1094 FL-----PAPEYVNO--LMPKKPS-----TAMVONOYNNISL 1124
QY 1186 -----AFGGAIVENPEYLTPOGGAAPPHPPPPAFSPAFDNLVYWDQ----- 1225
DB 1125 TATSKLPMDSRYQNSHSTAVDNPEYL-----NTNQSPKLVTFVSESPYIQSGNHQI 1176
QY 1226 --DPPE-----RGAPPSTFKGTPTAENPEYLGLDVP 1254
DB 1177 NLNDPDYQODFLPNETKPNGLLKVPAAENPEYLRVAAP 1214

RESULT 7
A47253
epidermal growth factor receptor, HER4 - human
C;Species: Homo sapiens (man)
C;Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 18-Jun-1999
C;Accession: A47253
R;Plozman, G.D.; Culouscou, J.M.; Whitney, G.S.; Green, J.M.; Carlton, G.W.; Foy, L.; N.
Proc. Natl. Acad. Sci. U.S.A. 90, 1746-1750, 1993
A;Title: Ligand-specific activation of HER4/p180erbB4, a fourth member of the epidermal
A;Reference number: A47253; MUID:93189574; PMID:8383326
A;Accession: A47253
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: nucleic acid
A;Residues: 1-1308 <PLO>
A;Cross-references: GB:L07868; NID:G337359; PIDN:AAB59446.1; PID:G337360
A;Note: sequence extracted from NCBI backbone (NCBI.P.126842)
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; growth factor receptor
F;716-981/Domain: protein kinase homology <KIN>
F;724-732/Region: protein kinase ATP-binding motif

[illegible]

Db	1009	DEEDLEMDMAEYLPV-QAFNIPPP	1056
Qy	1070	APRS-	1100
Db	1057	TPMSGNQFYVRDGGFAAEQGVSVYRAPVTSTIEAPVAQAGATAEIPDDSCCNGTLRKPAV	1116
Qy	1101	LPTHDPSPLQRYSEDPTVPLPS	1153
Db	1117	PHVOEDSSTQRYSDPTVFAPERSPRGELDEEGYTMPMRDKPQEYLNVPVE	1167
Qy	1154	EGPLPAARPAGATILERAKTLSPGKNGVVKDVFAGGAVENPEYLTQGGGAAPQHPHPPA-	1212
Db	1168	ENPFVSR-	1199
Qy	1213	-	1237
Db	1200	DEYVNEPLYLNTFANTLGKAYLKNLILSMPEKAKAFDNPDYWNHSLPFRSTLQHPDYL	1259
Qy	1238	KGTPT-	1249
Db	1260	QEYSTKYFKONGKIRPIVAENPEYL	1285
RESULT 8			
S06142			
Protein-tyrosine kinase (EC 2.7.1.112) mrk-Y precursor - southern platyfish			
N;Alternate names: epidermal growth factor receptor homolog; kinase-related tra			
C;Species: Xiphophorus maculatus (southern platyfish)			
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 18-Feb-2000			
C;Accession: S06142; S13809			
R;Wittbrodt, J.; Adam, D.; Malitschek, B.; Maeueller, W.; Raulf, F.; Telling, A.			
Nature 341, 415-421, 1989			
A;Title: Novel putative receptor tyrosine kinase encoded by the melanoma-induc			
A;Reference number: S06142; MUID:90015140; PMID:2797166			
A;Accession: S06142			
A;Molecule type: DNA			
A;Residues: 1-1166 <MIT>			
A;Cross-references: EMBL:X16891; NID:G65290; PIDN:CAA34770.1; PID:G65291			
R;Adam, D.; Maeueller, W.; Scharf, M.			
Oncogene 6, 73-80, 1991			
A;Title: Transcriptional activation of the melanoma inducing Xmrk oncogene in X			
A;Reference number: S13807; MUID:91125882; PMID:1846957			
A;Accession: S13809			
A;Status: preliminary; translation not shown			
A;Molecule type: DNA			
A;Residues: 821-1025, 'N', 1027-1098, 'A', 1100-1166 <ADA>			
A;Cross-references: EMBL:X56319; NID:G65284; PIDN:CAA39763.1; PID:G65285			
C;Genetics:			
A;Gene: mrk			
A;Map position: Y			
A;Introns: 872/3; 898/1; 947/1; 979/3; 1025/3; 1056/1			
C;Superfamily: epidermal growth factor receptor; protein kinase homology			
C;Keywords: ATP; growth factor receptor; phosphotransferase; transmembrane prot			
F;1-25/Domain: signal sequence #status predicted <SIG>			
F;26-1166/Product: kinase-related transforming protein (Tu) #status predicted <			
F;707-972/Domain: protein kinase homology <KIN>			
F;715-723/Region: protein kinase ATP-binding motif			
Query Match 39.2%; Score 2669; DB 1; Length 1166;			
Best Local Similarity 45.1%; Pred. No. 1e-101;			
Matches 572; Conservative 167; Mismatches 389; Indels 140; Gaps 30;			
Qy	4	AALCEWGLLLALLPPGAAS	59
Db	8	AALLQ--LLLVLSIRCCSDPDPRKVCQGTNSQNTM---	62
Qy	60	LELYLPTNASLSLQDIOEQVGVLLIAHQVROVPLQRLRIVRGQTQIFEDNYALAVLDN	119
Db	63	LEIYITQENQDLSFLQSIQEVGGVLLAMNEVSTIPLVNLRLIRGQNLVEGNFTLLVMSN	122
Qy	120	GDPLNNTVPTGASPGGLRLQSLRSLTEILKGGVLIQRNPOLCYQDTILWKDIPHKNNQL	179

RESULT

S06142

protein-tyrosine kinase (EC 2.7.1.112) mrk-y precursor - southern platyfish
N;Alternate names: epidermal growth factor receptor homolog; kinase-related tra
C;Species: Xiphophorus maculatus (southern platyfish)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 18-Feb-2000
C;Accession: S06142; S13809
R;Wittbrodt, J.; Adam, D.; Malitschek, B.; Maeueller, W.; Raulf, F.; Telling, A.
Nature 341, 415-421, 1989
A;Title: Novel putative receptor tyrosine kinase encoded by the melanoma-inducin
A;Reference number: S06142; MUID:50015140; PMID:2797166
A;Accession: S06142
A;Molecule type: DNA
A;Residues: 1-1166 <WIT>
A;Cross-references: EMBL:X16891; NID:g65290; PIDN:CAA34770.1; PID:g65291
R;Adam, D.; Maeueller, W.; Scharf, M.
Oncogene 6, 73-80, 1991
A;Title: Transcriptional activation of the melanoma inducing Xmrk oncogene in X
A;Reference number: S13807; MUID:91125882; PMID:1846957
A;Accession: S13809
A;Status: preliminary; translation not shown
A;Molecule type: DNA
A;Residues: 821-1025 'N',1027-1098, 'A',1100-1166 <ADA>
A;Cross-references: EMBL:X56319; NID:g65284; PIDN:CAA39763.1; PID:g65285
C;Genetics:
A;Gene: mrk
A;Map position: Y
A;Intons: 872/3; 898/1; 947/1; 979/3; 1025/3; 1056/1
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; growth factor receptor; phosphotransferase; transmembrane prot
F;1-25/Domain: signal sequence #status predicted <SIG>
F;26-1166/Product: kinase-related transforming protein (Tu) #status predicted <
F;707-972/Domain: protein kinase homology <KIN>
F;715-723/Region: protein kinase ATP-binding motif
Query Match 39.2%; Score 2669; DB 1; Length 1166;
Best Local Similarity 45.1%; Pred. No. 1e-101;
Matches 572; Conservative 167; Mismatches 389; Indels 140; Gaps 30;
Qy 4 AALCRWGLLLAALLPPGAAST-----QVCTGDMKRLRPASPTHLDMLRHLYQGCVVQGN 59
Db 8 AALLQ--LLLVLSIRCCSTDPRKVCQGTSNQWTM---LDNHYLKMVKMYSGCNVYLEN 62
Qy 60 LELYLPTNASLSFDIQIEQGVYVLIHNOVRQPLQRILRVRGTLFDNNVALAVLDN 119
Db 63 LEITYTQENQDLSFQSQIEQGVGVLIAMNEVSTIPLVNRLRLRGONLYEGNFTLLVMSN 122
Qy 120 GDPNNNTTVPVTCSPGGLRLQSLRSLTEILKGGVLIQRNPOLCYQDFTILWKDIFHKNNQL 179

Db 411 VFSNLTTIGRSLYNRGFSLLIMKNLNTVSLGFRSLKEISAGRIYISANRQLCVHSLNW 470
Qy 483 DQIFRNPHQALLHTA-NRDECEVGEGLACHQLCARGHCWPGPTQCVNCSQFLRGQECV 541
Db 471 TKVLRGPTBERLDIKHNRPRDCVAEGKVCPLDCLSSGGCGWPGGQCLSCRNYSRGVCV 530
Qy 542 EECRVGLGLPREVYNARHCLPCHPECOPOGNSVTCFGEADOCVACAHYKDDPPFCVARCP 601
Db 531 THCNFLNGEPREFAHEAECSCHPECOPEMGATCNGSGDTCACQAHFRDGHCVSSCP 590
Qy 602 SGVKPDLISYMPIWKFPPDEAGACQPCINCTHSCVDLDKGCAPBQRA-----SPLTSIVSA 657
Db 591 HGVLG--AKGPIYKVPQVNECRCHENCCTGCKGPELQDCLGQTLVLIGKTHLTWALT 648
Qy 658 VVGILLVVLGVFGIILKRQKIR-KYWRRLLOETELVEPLTPSGAMPNQAMRILK 716
Db 649 IAG--LVVIFMMLGGTFLYMRGRIONKRAMRRYLERGESIEPLDPS-EKANKVLARIFK 705
Qy 717 ETELKRVKVLGSGAGFTVYGIWIPDGENVKIPVAIKVLRENTSPKANKILDEAYVMAG 776
Db 706 ETELKRVKVLGSGVFGVHKGVWIPEGESIKIPVCIKVIDKSGRQSFQAVTHMLAIGS 765
Qy 777 VGSYVSRLLGICLTSTVQLVTLQMPYGCLLDHVRENRLGSLQDLLNWCQIAKMSYL 836
Db 766 LDHAHIVRLGLCPGSSQLQVLYPLGLSLLDHVRQHRGALGPQLLLNMGVQIAKMYL 825
Qy 837 EDVLRVHDLAARNVLKSPNHVKITDFGLARLLDIDETEHADCGKVPKIMWALESLR 896
Db 826 BEHGMVHRNLAARNVLKSPSQVADFGVADLLPPDDKQLLYSEAKTPIKMWALESIHF 885
Qy 897 RRFTHQSDVWSYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPQPTCTIDVYMIW 956
Db 886 KGYTHQSDVWSYGVTVWELMTFGAEPTAGLAEVPLLEKGERLAQPTCTIDVYMWV 945
Qy 957 KCMWIDSECRPRFRELSEFSRMARDPQRFVVIQNEDLGPA---SPLDSTFYRSLLEDD 1013
Db 946 KCMWIDENIRPTEKLANEFTRMARDPPRYLVIKRES-GPGIAPGPEPHGLTNKKLEVE 1004
Qy 1014 MGDVDAEYLVPOQFCFDDPPAGAGVWHRHRSSTSGGDLTLGLEP-SEEAAPR 1072
Db 1005 LEPELDLDLLEAED-----NLATTLGSLSLPVTGLNRRGSQ 1045
Qy 1073 SPLAPSGAGSDVFDGLMGAAKGLQSLPTH--PSPLQRYSEDPVPLP-----SETD 1125
Db 1046 SLLSPSSGY-MPMNQNLGESCQESAVSGSERCPRVSLH-----PMRGCLASESE 1098
Qy 1126 GYVA-----PLTCSQPE-----YVNPQDVRPQPPSPREGP----- 1156
Db 1099 GHVTGSEAEQLQKVMCRSRSRSPRGRDSAVHSQRHSLLTPTVPLSPGLEEDVNG 1158
Qy 1157 --LPAARPAGATLERAKTLSP-GKNGV-----KDVAFGCAVENPEYLTPOGGAAPQ 1207
Db 1159 YVMPDTHLKGTPSSREGTLSSVGLSGVLGTEEDED-----EYEYNNRRRHSF-P 1209
Qy 1208 HPPAFSPAFDNLTYWD-----QDPRERGAPPSTFKCTPTAENPEYL 1249
Db 1210 HPPRPSLBEELGYEYMDVGSLSASLGSTQSCPLHPVIMPTAGTTTDEDEYEM 1263

RESULT 10
JC4387
epidermal growth factor receptor homolog precursor - rat
N;Alternate names: ErbB3 protein; HER3 protein
C;Species: Rattus norvegicus (Norway rat)
C;Date: 17-Jan-1996 #sequence_revision 19-Apr-1996 #text_change 13-Nov-1998
C;Accession: JC4387
R;Hellver, N.J.; Kim, H.H.; Greaves, C.H.; Sierke, S.L.; Koland, J.G.
Gene 165, 279-284, 1995
A;Title: Cloning of the rat ErbB3 cDNA and characterization of the recombinant protein.
A;Reference number: JC4387; MUID:96096535; PMID:8522190
A;Accession: JC4387
A;Molecule type: mRNA
A;Residues: 1-1339 <HEL>

A;Cross-references: GB:U29339; NID:G915389; PID:G915390
A;Experimental source: liver
A;Note: The authors translated the codon AAC for residue 369 as Thr and GRT for residue 370
C;Comment: This protein is a functional heregulin receptor that transduces signals to the
C;Genetics:
A;Gene: ErbB3
C;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homol.
C;Keywords: ATP; growth factor receptor; liver; phosphoprotein; transmembrane protein
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-1339/Product: epidermal growth factor homolog #status predicted <MAT>
F;640-659/Domain: transmembrane #status predicted <TM>
F;705-970/Domain: protein kinase homology <KIN>
F;713-721/Region: protein kinase ATP-binding motif
F;939,1051,1156,1194,1196,1219,1257,1259,1273,1286,1325/Binding site: phosphate (Tyr)
Query Match 34.2%; Score 2328.5; DB 2; Length 1339;
Best Local Similarity 40.6%; Pred. No. 8.5e-88;
Matches 522; Conservative 171; Mismatches 433; Indels 159; Gaps 35;
Qy 3 LAALCRWGLLLALLPPGAA---STQVCTGTDKMLRLPASPTHLDMLRHLYQGCQVVQGN 59
Db 7 LQVLC-----FLLSLARGSEMGNSQAVCPGTLNGLSVTGADNQYQYLYKEVGVWGN 62
Qy 60 LELTYLPTNASLFLQDIOEVQGYLIAHQVRQVPLQRLRIYVRGTQLFEDNVALAVLDN 119
Db 63 LEIVLTGHNADLSFLQWIREVTAYVLVAMNEFSVLPLNLRVVRGTVQVYDGEKFAFV-- 120
Qy 120 GDPINNTPTVGTASPGGLRELQRLSLEILKGVLTQNPOLCYQDTILWKDIFKHNQL 179
Db 121 ---LNTNT-----NSSHALRQLKFTQLTEILSGGVYIEKNDKLCMTDIDRDRVR-- 170
Qy 180 ALTLIDNRSRACHPCSPCKSGSRGSESSDCQSLTRTVACAGC--ARCKGLPTDCCHE 238
Db 171 GAEIVKNGANGCPPCHEVCKG--RCWGPDPDCCQILTKLICAPQNGRCFEPNQCCHD 239
Qy 239 QCAAGCTGPKHSDCLACLHFNHSGICELCPALVYNTDTFESMPNPEGRYTEGASCVTA 298
Db 230 ECAGGCGPQDTCFACRFENDSGACVPRCPPELVYNKLTFFOLEPNPHTKYGVGVAS 289
Qy 299 CPVNYLSTDVGSCTLVCPPLHNQVTAEDGTQCEKSKPCARVCYGL--GMQYIKANSKF 356
Db 290 CPNFV--VDQTFVCRACPPDKMEVD--KHGLMKMCEGGLCPKACEGTGSGSRVYQTVDSN 347
Qy 357 IGITELEFAGCKKIFGSLAPLPESFDGDPASNTAPLQPEOLQVFETLEETGYLYTSAMP 416
Db 348 ID-----GFVNCTKILGNLDELITGLNVDPWHKIPALDPEKLVNFRVREITGYLNTSWP 403
Qy 417 DSLPDLISVFONQVIRGRILHNGAYS--LTQLGLGISWGLRSLRSLRELGSGLALIHNTSLC 475
Db 404 PHMHFVSFNLTTIGRSLYNRGFSLLIMKNLNTVSLGFRSLKEISAGRIYISANQQLC 463
Qy 476 FVHTVPWDQLFRNPHQALLHTA-NRDECEVGEGLACHQLCARGHCWPGPTQCVNCSQF 534
Db 464 YHLSLWNLRLRGSPBERLDIKYDRPLGECLAEGKVCPLDCLSSGGCGWPGGQCLSCRN 523
Qy 535 LRQCEVCEECRVLOGLPREVYNARHCLPCHPECOPOGNSVTCFGEADOCVACAHYKDDP 594
Db 524 SRGVCVTHCNFQGEPEFVHEAQCFSCHPECLPMEGTSTYNGSGSDACARCAHFRDGP 583
Qy 595 FCVARCPGKVPDLISYMPIWKFPPDEAGACQPCINCTHSC--VDLDDKGCAPBQASPLT 652
Db 584 HCYNSCPFGILG--AKGPIYKVPDAQNECRCHENCCTGCGNPELQDCLQAEVLSMKPH 641
Qy 653 SIYSAVVIGILLVVLGVFGIILKRQKIR-KYWRRLLOETELVEPLTPSGAMPNQAO 711
Db 642 LVIAVTVG--LAVILMLGGSFLYMRGRIONKRAMRRYLERGESIEPLDPS-EKANKVL 698
Qy 712 MRILKETELRKVKVLGSGAGFTVYGIWIPDGENVKIPVAIKVLRENTSPKANKILDEA 771
Db 699 ARIFKETELRKVKVLGSGVFGVHKGVWIPEGESIKIPVCIKVIDKSGRQSFQAVTDM 758
Qy 772 YVMAGVGSYPVSRLLGICLTSTVQLVTLQMPYGCLLDHVRENRLGSLQDLLNWCQIAK 831

Db 759 LAVGSLDHAHIVRLGLCPGSSQLVTOYLPLGSLLDHVKHRETLPQQLLWNGVOIAK 818
Qy 832 GMSYLEDLVRLVHRLDAAARNVLKSPNHVKITDFGLARLLDIDETEYHADGGKVPKIMWAL 891
Db 819 GMYLEHSMVHRDLALRNWMLKSPSQVQVADFGVADLLPDDKQLLHSEAKTPIKIMWAL 878
Qy 892 ESILRRPFTHOSDVMSYGVTVWELMTGAKPYDGIPIAREIPDLLEKGERLPPOPICTIDV 951
Db 879 ESIHFGKYTHOSDVMSYGVTVWELMTGAEYAGILRLAEIPDLLEKGERLAQPOICTIDV 938
Qy 952 YIMVVKWMDISECRPRFRELVSFSEFARMARDPQRFVVIQNEIDLPASPLDSTFVRSLLD 1011
Db 939 YIMVVKWMDISECRPRFRELVSFSEFARMARDPQRFVVIQNEIDLPASPLDSTFVRSLLD 995
Qy 1012 DMDGDLVDAEYLVPOQGFPCPDAPGAGGMVHRRHSSSTRSGGDLTLGLSESEF--- 1068
Db 996 KEL-----QEALEPEL-----DLDLDLEAEERGLA 1021
Qy 1069 -----EAPRSLAPSEG-----AGDVDFGDLGMGAAGLQSLPT 1103
Db 1022 TSLGSLSLPTGTLTRPRGSQLSPSSGYMPNMNQSSLGAEACLSAVLGGREQFSRDISL 1081
Qy 1104 HDPSPLQKYSDDPTVPLPSETDGVV---APL-----TC-----SPOPE---VYNOP 1143
Db 1082 H-PIPRG-----PASESEGHVTCSEAELOKVKVCSRSRSPRFRGDSAYHSQR 1133
Qy 1144 DVRRQPPSPREGP-----LPAAPAGATLERAKTLSP-GKNGV-----KDVF 1185
Db 1134 HSLITPTVPLSPGLEEDGNGVYMPDTHLRGSSSREGTLSSVGLSSVLGTEBEDED-- 1191
Qy 1186 AFGAVENPEYLTPOGGAAAPQHP 1210
Db 1192 -----BEYEMNKRKRGSP-PRPP 1209

RESULT 11
TVFVLV
N;Contains: amino end of gag protein; env protein fragment; protein-tyrosine kinase
C;Species: avian leukosis virus, ALV
C;Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 11-Jun-1999
C;Accession: B00643
R;Nilsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M.
Cell 41, 719-726, 1985
A;Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and product formation
A;Reference number: A00643; MUID:85228222; PMID:2988784
A;Accession: B00643
A;Molecule type: mRNA
A;Residues: 1-698 <N1>
A;Cross-references: GB:M10066; GB:M13881; NID:G211749; PIDN:AAA48763.1; PID:G211750
A;Note: in Genbank entry CHKERBBF, release 109.0, the source is designated as Gallus gallus
C;Comment: This protein is synthesized as a gag-env-erbB protein.
C;Genetics:
A;Gene: gag-env-erbB
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific phosphatase
F;1-6/Product: gag protein (fragment) #status predicted <CAG>
F;7-59/Product: env protein (fragment) #status predicted <ENV>
F;60-698/Product: protein-tyrosine kinase erbB #status predicted <ERB>
F;194-459/Domain: protein kinase homology <KIN>
F;202-210/Region: protein kinase ATP-binding motif
F;229/Active site: Lys #status predicted

Query Match 25.9%; Score 1766.5; DB 1; Length 698;
Best Local Similarity 52.28; Pred. No. 3.5e-65;
Matches 374; Conservative 80; Mismatches 137; Indels 125; Gaps 18;

Qy 578 GPEADQCVACAHYKDPFCVVARCPGSKVPLSYMIWKFDEGACQPCPINCTHSCVDL 637
Db 60 GP--DHCMCAHFIDGPHCVKACPAAGVLGENDTL-VMKYADANAVCOLCHPNCRTCKGP 116
Qy 638 DDKGCPAEQASPLTSSIVAVV-GILLVVVLGVVFGILIKRQOKTKYTMRLLOTEL 696

Db 117 GLESCP---NGSKTPSIAAGLVGVGLLCLVVVGIGIGLYLRRR-HIVKRTLRRLQREL 172
Qy 697 VEPLTPSGAMPNOAMRIILKETELRKVKVLGSAFGTVYKGIWIPDGENVKIPIVAIKVL 756
Db 173 VEPLTPSGEAPNAHLRIILKETEFKKVKVLGSAFGTVYKGLWIPGEKVKIPIVAIKEL 232
Qy 757 ENTSPKANKEILDEAYVMAGVGSPPYVSRLLGICLTSTVQLVTOLMPYGCILLDHVRENRR 816
Db 233 EATSPKANKEILDEAYVMASVDNPHVCRLLGICLTSTVQLITOLMPYGCILLDIREHKN 292
Qy 817 LGSODLLNMCWOTAKMSVLELDVRLVHRLDAAARNVLKSPNHVKITDFGLARLLDIDETE 876
Db 293 IGSQYLLNMCVQITAKGNVLEERLVRDLAARNVLKTPQHVKITDFGLAKLGLGADEKE 352
Qy 877 YHADGGKVPKIMWALSIILRRRTHOSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLE 936
Db 353 YHAEGGKVPKIMWALSIILHRIYTHOSDVMSYGVTVWELMTFGSKPYDGIPIASEISSVLE 412
Qy 937 KGERLPPOPICTIDVIMVVKWMDISECRPRFRELVSFSEFARMARDPQRFVVIQ-NEDLG 995
Db 413 KGERLPPOPICTIDVIMVVKWMDISADSRPKFRELIAEFSKWARDPPRYLVITQGERMH 472
Qy 996 PASPLDSTFVRSLLDMDGDLVDAEYLVPOQGFPCPDAPGAGGMVHRRHSSSTRSG 1055
Db 473 LPSPTDSKFTVRLMEBEDMEDIVDAEYLVPHQGF-----NSPST--- 513
Qy 1056 GGDLTGLGPSEBEAPRSL-----APSEGAGSDVDFDGLGMGAAGLQSLPTHDPSPLO 1110
Db 514 -----SRTPLLSLSLTSNNSAINCID-----RNGQHPVRESFVQ 550
Qy 1111 RYSEDPVPLPSET--DGVVAPLTCSPQPEYVNPQDVRPQPPSPREGPLPAARPAATLE 1168
Db 551 RYSSDPTGNFLESIIDGFL-----PAPEYVNVQ--LMPKKPS----- 585
Qy 1169 RAKTLPSPKNGVVKDVF-----AFGAVENPEYLTPOGGAAAPQHPPPAP 1213
Db 586 -----TAMVQNIYNNISLTAISKLPMDSRYSQNSHSTAVDNPEYL-----NTNQSLA 633
Qy 1214 SPAFDNLVYWDQ-----DPE-----RGAPPSTFGTPTAENPEYVGLDVP 1254
Db 634 KTVFESSPWIQSGNHQINLDPDYQODFLPNETKNGLLKVPAAENPEYLVRAAP 689

RESULT 12
TVVUH
protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus (strain H)
C;Species: avian erythroblastosis virus
C;Date: 18-Apr-1984 #sequence_revision 18-Apr-1984 #text_change 11-Jun-1999
C;Accession: A00644; A38022
R;Yamamoto, T.; Nishida, T.; Miyajima, N.; Kawai, S.; Ooi, T.; Toyoshima, K.
Cell 35, 71-78, 1983
A;Title: The erbB gene of avian erythroblastosis virus is a member of the src gene family
A;Reference number: A00644; MUID:84026539; PMID:6313229
A;Accession: A00644
A;Molecule type: DNA
A;Residues: 1-604 <X>
R;Debuire, B.; Henry, C.; Benaissa, M.; Biserte, G.; Claverie, J.M.; Saule, S.; Martin, Science 224, 1456-1459, 1984
A;Title: Sequencing the erbA gene of avian erythroblastosis virus reveals a new type of ATP-binding motif
A;Reference number: A38022; MUID:84223957; PMID:6328658
A;Accession: A38022
A;Molecule type: DNA
A;Residues: 1-28, 'W', 30-139, 'P', 141-145, 'V', 147-152 <DEB>
A;Cross-references: GB:K02006
C;Genetics:
A;Gene: erbB
C;Superfamily: epidermal growth factor receptor; protein kinase homology
C;Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific phosphatase
F;130-395/Domain: protein kinase homology <KIN>
F;138-146/Region: protein kinase ATP-binding motif
F;165/Active site: Lys #status predicted

Query Match	24.2%	Score 1647;	DB 2;	Length 544;
Best Local Similarity	54.9%;	Pred. No. 2e-60;		
Matches 345;	Conservative 70;	Mismatches 121;	Indels 92;	Gaps 15
578 QY	GPEADQCACAHYKDPFCVCARPCSGVKPDLSPMIPIKFPDEEGACQCPINCTHSCVDL	637		
1 Db	1 GP--DHCMCKAHFDIDGPHCVKACPAVGLENDTL-VWKYADANAVCQLCHPNCTRGCKGP	57		
638 QY	DDKGCPEAQERASPLTISIVSAV--GILLVVLGVVFGILIKRRQKIRKYTRRLLOETEL	696		
58 Db	GLECCP---NGSKTPTSAAGVUGLLCLVVGLGIGLYLRRR-HIVRKFTLRLLQEREL	113		
697 QY	VEPTTPSGAMPQAOQRILKETELRKVKVLGSGAFGTVYKGIWIDGENVKIPIVAIKVL	756		
114 Db	VEPTTPSGEAPNOAHLRIKETEFKKVKVGLGFAFGTIVYKGLWIEGEKVTIPVAIKEL	173		
757 QY	ENTSPKANKELIDEAAYVMAGVGSPVYSRLIGICTSTVOLVTQLMPYGCLLDHHVRENGR	816		
174 Db	EATSPKANKELIDEAAYVMASVDNPHVCRLLGICLSTVQLITQLMPYGCLLDIYIREKDN	233		
817 QY	LGSODLLNWCQIAKGSYLEDVRLVHRDLAARNVLKVS PNHVKITD FGLARLLDIDETE	876		
234 Db	IGSQVLLNWCQIAKGMNYLEERHVRDLAARNVLKVPQHVKITDFGLAKQLGADEKE	293		
877 QY	YHADGGKVPDKWALESILARRTHOSDVWSYGVTVWELMTGAKPYDGI PAREIPDILLE	936		
294 Db	YHABGGKVPDKWALESILHRIYTHOSDVMSYGVTVWELMTGSKPYDGI PASEISSVLE	353		
937 QY	KGERLPOPPICTIDVYIMVYKMWIDSECRPFRELVSFSFMSWARDPQRFVVIQ--NEDLG	995		
354 Db	KGERLPOPPICTIDVYIMVYKMWSDASRPKRELIAEFSKWARDPPRYLVIQGERMH	413		
996 QY	PASPLDSTFYRSLLEDMDGLVDABEYLPVQQGFPCPDPAFCAGGMVHHRHSSTRSRG	1055		
414 Db	LPSPTDSKFYRTLMEEDMEDIVDABEYLVPHQGF-----NSPST---454			
1056 QY	GGDTLTGLEPSEEA PRSL-----APSEGAGSDVFDGDLGMAKGLQSLTHDPSPLO	1111		
455 Db	-----STPLLSLSLATSNNSATNCIDRNGG-----H-----481			
1111 QY	RYSEDPVPLPSETDGVVAPLTSCPOPEYVQPDVRAPOQPSPREGPLPAARPAAGAT--LER	1165		
482 Db	-----PVREDGFL-----PAPEYVQQ--LMPKFPSTAMVQVQNIYVILSTAIKSK	523		
1170 QY	AKTLSPGKNGVYKDVFAFGAVENPEYL	1197		
524 Db	LPIDSRYON-----SHSTAVDNPEYL	544		

REGOUT
S00727

RESULT 13
S35745
protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus
C:Species: avian erythroblastosis virus
C:Date: 03-Mar-1994 #sequence_revision 26-May-1995 #text_change 28-Feb-1997
C:Accession: S35745
R:Vennstroem, B.
submitted to the EMBL Data Library, March 1993
A:Reference number: S35743
A:Accession: S35745
A:Molecule type: DNA
A:Residues: 1-544 <VEN>
A:Cross-references: EMBL:X12707
C:Genetics:
A:Gene: erbB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: Arr; oncogene; phosphotransferase; transforming protein; tyrosine-specific H
F:135-400/Domain: protein kinase homology <KIN>
F:143-151/Region: protein kinase ATP-binding motif
F:170/Active site: Lys #status predicted

Query Match	24.1%	Score 1640;	DB 2;	Length 545;
Best Local Similarity	54.9%	Pred. No. 3.9e-60;		


```
Qy 720 LRKVVILGSAFGTYVYKGIWIPDGENVKIPVAIKVLENTSPKANKEILDEAYUMAGVGS 779
Db 810 LRKGGVLGMGAFGRVYKGVWPEGENVKIPVAIKELKSTGABSESEFLREAYIMASEEH 869
Qy 780 PYVSRLLGICLTSTVOLVTOIMPYGCLLDHVRENRGRGLSGODLNNWCMQIAKMSYLEDV 839
Db 870 VNLKLLAVCMSSOMLITOLMPLGCLLDYVRNRDKIGSKALLNWSQIAKMSYLEEK 929
Qy 840 RLVRDLAARNVLVK---SPNHVKITDGLARLLDIDETEHADGGKVPKIMMALESILR 896
Db 930 RLVRDLAARNVLVRLLAGEDH---DFGLAKLSSDSNEYKAAGGKMPKWLALECIRN 985
Qy 897 RRFTHQSDVMSYGVTVVHMLTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYIMV 956
Db 986 RVFTSKSDVMAFGVTIWEILLTFQORPHENIPAKDIPDLIEVGLKLEQPEICSLDIYCTLL 1045
Qy 957 KCWMIDSECRPRPRELVSERFMRMARDPQRFVVIQNEIDLG--PASPLDSTFYRSLLEDD-- 1012
Db 1046 SCWHLDAAMRPTFKQLTTVFAEFARDPGRYLAILGDKFRLPA-----YTSQDEKDLI 1098
Qy 1013 -DMGDLVDAEYLVPOQGFPCDPAPGAGGMVHHRSSSTRSGGDLTLGLEPSEEEAP 1071
Db 1099 RKLAPTTDGSIAIAKPDYLOPKAALGPS-----HRTDCT-----DEMP 1137
Qy 1072 -----RSPLAPSEGAGSDVFDG---DLGMAAKGLQSLPTHDPSPLOQRYSEDTVPLPS 1122
Db 1138 KLNRYCKDPSNKNSTGDDERDSSAREVGVGNLR-----LDLPV 1176
Qy 1123 ETDGYVAPLTCSPQPEYVNOQDVPRPQPPSPREGPLPAARPAGATLERAKTILSPGKNGVVK 1182
Db 1177 DEDDYLMPTCQPGFNNNNNNNN-----NPNQNNMAAVGVAAGYM----- 1214
Qy 1183 DVFAFGGAVENPEYL---TPQGAAPQPH-----PPPAFSP-AFDN 1219
Db 1215 DLIGVPVSDNPEYLLNAQTLGVGESPIPTQITIGIPVNGGPGTMEVKVPMGPGSEPTSSDH 1274
Qy 1220 LYWD 1224
Db 1275 EYND 1279
```

Search completed: July 22, 2003, 09:09:23
Job time : 30.9062 secs